5-2017

A Novel Method for Quantifying Spatial Patterns in Plants

Logan H. George
College of William and Mary

A Novel Method for Quantifying Spatial Patterns in Plants

A thesis submitted in partial fulfillment of the requirement
for the degree of Bachelor of Science in Biology from
The College of William and Mary

By

Logan Hayden George

Accepted for Honors

Dr. Josh Puzey, Director

Dr. Helen Murphy

Dr. Matthias Leu

Dr. Jordan Walk

Williamsburg, VA
May 3, 2017

A Novel Method for Quantifying Spatial Patterns in Plants

Logan George

Defense Date: May 3rd at 9:00am in ISC 1111

Dr. Joshua Puzey, Advisor

Honors Thesis
Degree of Bachelor in Science with Honors
In Biology

College of William & Mary

Williamsburg, Virginia

Abstract

Color patterns are found in a plethora of organisms, from vertebrates to flowering plants. While many studies have examined the mechanisms that produce these diverse patterns in animals, little research has investigated the mechanisms by which plants create color patterns. The conclusions drawn from animal studies may not accurately translate to plants due to early divergence in the evolution of life. Characterization of plant patterning mechanisms would have widespread impacts on developmental and evolutionary biology. To unravel the mystery behind pattern formation, we suggest an experimental framework to understand pattern evolution and development at a phenotypic, genotypic, and quantitative level, creating a holistic model for the evolution of complex traits and phenotypic diversity. Here, we provide a novel protocol for the quantification of pattern morphology, and demonstrate its efficacy in a segregating F2 population of the model organism *Mimulus luteus*. By co-opting ArcGIS and FragStats, two landscape ecology softwares, to map petal patterns, we developed a high throughput method for objective phenotype characterization. This protocol is useful for preliminary work in a bulk segregant analysis by separating a population in discrete groups based on morphology. We used this protocol to demonstrate that patterns are distinct between petals within the same flower depending on petal location, and that there is a genetic basis for pattern formation in flowers. Minor tweaks to the genes guiding pattern formation may be responsible for the rapid evolution of angiosperm flower diversity. Future work is required to identify the genes responsible for pattern formation, and to develop a method for modeling these genes to predict how minor mutations would impact phenotypic traits.

Introduction

Pattern Formation in Organisms

A pattern is defined as a repeated form or design that is recurring, and is controlled during its creation (*Merriam-Webster's Collegiate Dictionary*, 1999). One can identify a pattern by eye alone based on its regularity and intelligibility, making patterns distinct from random organization. In nature, color patterns have evolved in a myriad of organisms (Fig. 1). Is this diversity generated at random? This does not appear to be the case, based on prior work that has characterized the regularity of patterns in several organisms (See: Nadeau and Sack, 2002; Kunte et al., 2014; Mallarino et al., 2016).

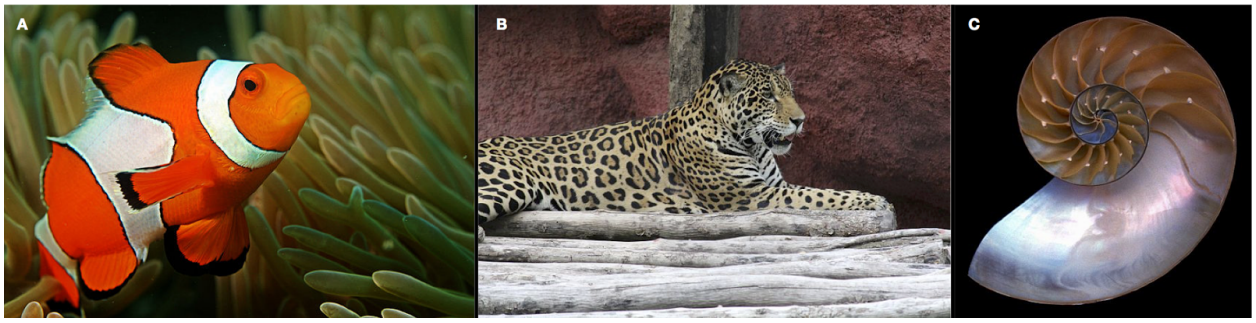

Figure 1: Color patterns in nature are very diverse across life. **A**: The orange-and-white striping characteristic of the clownfish. **B**: Leopards develop spots to enhance camouflage. **C**: Nautilus shells display a regular logarithmic spiral, seen by cross-section. Open sources images from Wikimedia Commons users Ritiks, Aradhanait, and Sergio Valle Duarte.

Therefore, a fundamental question emerges: how are these patterns created at a developmental and genetic level? It would appear that these patterns do not develop randomly, but rather through some manner of genetic regulation. Development is a stringently controlled process mediated by temporal and spatial gene expression that guides overall organization, such

as neuronal connection formation during corticogenesis (Greer and Greenburg, 2008), logarithmic spiraling in nautilus shells (Ball, 2013), and phyllotaxis in plant leaf organization (Reinhardt et al., 2003). All of these complex patterns likely follow discrete genetic rules. Because patterns are formed through regulated pathways, it might be assumed that small tweaks to these genetic rules would produce the breadth of pattern diversity seen in nature.

Patterns in the Animal Kingdom

Several natural patterning mechanisms have been well-characterized, and there are some examples that have pinned down a genetic basis to the formation of these patterns. One of these cases involves *Rhabdomys pumilio*, a species of African striped mice. Recent research has identified a single gene, *Alx3*, as the sole modulator of dorsal stripe development in these mice (Mallarino et al., 2016). *Alx3* creates these patterns by inhibiting *Mitf*, a gene that triggers melanocyte maturation. Here, *Alx3* defines the outer limit of "dark" during stripe development, creating sharply defined patterns on the dorsal side of the mouse. In a similar manner, research has demonstrated that butterflies of the *Papilio* genus create mimetic patterns through control of a single locus: *doublesex* (Kunte et al., 2014). *Dsx* controls a supergene, a tightly linked locus that guides the formation of the entire wing pattern in females. While previous hypotheses had believed that these complex patterns formed due to many tightly linked genes (Clarke and Sheppard, 1972), Kunte's study demonstrates that only a single gene is responsible for wing spot development.

Pattern Formation in Plants

It is important to note that multicellularity evolved multiple times throughout the history of life (Knoll, 2011). This fact suggests that plants and animals likely underwent different courses of evolution regarding many physiological and molecular mechanisms, including pattern formation. Due to the divergent life history of plants and animals, plants likely form patterns in a manner that is genetically distinct from animals. Thus far, all examples of pattern formation discussed here pertain to animal models, but the conclusions drawn from these studies may not accurately translate to plants. Despite the void in our knowledge of plant pattern formation, some studies have looked at the organization of plant development, specifically regarding spatial organization of stomata, the openings in leaves responsible for gas exchange. In plants, a pattern-control gene has been identified that coordinates the placement of stomata. Rather than being placed by random distribution, a single gene, *TMM*, or *Too Many Mouths*, modulates the spatial organization of stomata; loss of *TMM* leads to stomata overproduction and loss of organization (Nadeau and Sack, 2002).

This example, however, only discusses structural layout, and does not provide insight as to how color patterns are created. Color patterns are particularly prominent in flowering plants, or angiosperms. Angiosperms are a subkingdom of plants characterized by the evolution of a complex flower, separating this group from the flowerless gymnosperms. In angiosperms, there is a huge level of diversity between flower pattern designs (Fig. 2), but not much is known about how these designs are formed during development. Some recent work has investigated how color borders are defined in the petal during development. *LAR1* has been identified as a transcription factor that commits metabolites, specifically flavonols, to the pigment biosynthesis pathway

(Yuan et al., 2016). Through *LAR1* expression, flavonols are either unmodified, maintaining a light color, or converted into dark anthocyanin pigments to produce shades of red or purple. This work, while groundbreaking in the field of angiosperm developmental genetics, only reveals a piece of the grander puzzle regarding spatial organization of complex pigment patterns. We seek to elucidate how the most complex patterns arise in the most simplistic manner across angiosperms.

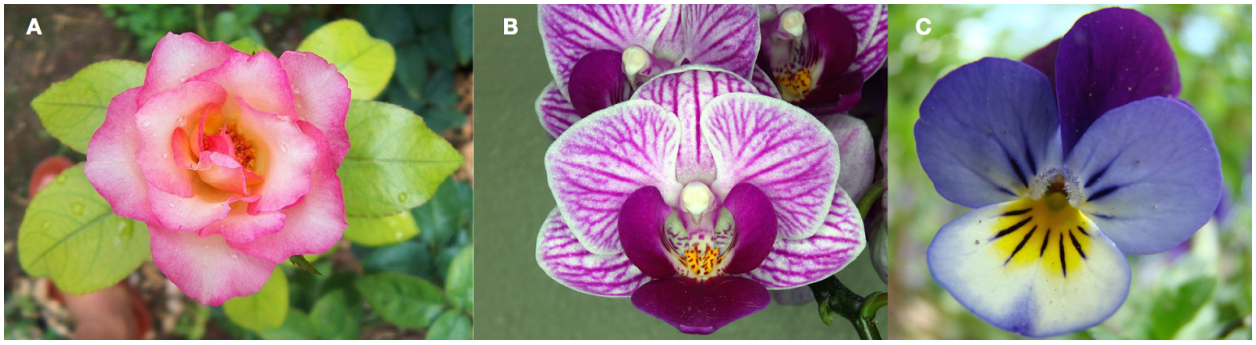

Figure 2: Flowering plants create diverse petal patterns during development. **A**: Roses can create a gradient of pigment. **B**: Orchids may form branching pigment "veins" using dark purple pigment. **C**: Pansies create a color gradient on their petals, with different areas of the petals expressing various pigments. Open source images from Wikimedia Commons users Agguizar, Tejvan Pettinger, and Wisnoiowy.

Figure 3: Flower pattern is diverse across closely related species. Flowers from the genus *Mimulus* show a widespread breadth of flower patterning. These patterns are distinguishable by color intensity, gradient formation, spottiness, and pigment aggregation.

Pattern Diversity in Angiosperms

Within angiosperms, there are estimated to be as many as 400,000 flowering plant species (Edwards, 2010), with many demonstrating radical variation in structure and morphology of the flower and its design. Even among individuals of the same genus, there can be dramatic variation in flower design (Fig. 3). Given that pollinators are often drawn to flowers based on pigment and flower design, it has been suggested that variation in flower morphology can lead to pollinator-dependent speciation events (Medel et al., 2003). As such, petal pattern variation is an area of interest in both evolutionary and developmental biology.

Because pattern variation has implications in several biological fields, it serves as a good model system for studying the genetics of development. In particular, angiosperms are an interesting area of study because they display a huge range of patterns, yet these plants have evolved in a relatively short time frame of approximately 150 million years (Bodt et al., 2005). Little work has characterized how these patterns evolved so rapidly. One simple possibility is that plants use a basic genetic framework to control pattern development, which is subject to small tweaks in the coding or regulatory domains. Through this framework, novel patterns could be formed by small changes in gene expression, at either the regulatory or effector level. To put it simply, this framework would be an easy mechanism to create novel patterns through the evolution and accumulation of minor genetic changes. For example, a transcription factor in one species that regulates pattern development may cause aggregated spots to form. A change in the promotor strength of this transcription factor, which could occur through single nucleotide polymorphisms (SNPs) and other random mutations, may result in a dispersed spot phenotype. However, the genes at most levels of spatial pattern formation in plants have not yet been

identified, so we seek to identify these genes in order to construct this framework for creating diversity.

Analyzing Petal Pattern Formation in Mimulus Luteus

To model petal pattern variation, we turn to *Mimulus*, an emerging model organism in the field of evolutionary genetics. *Mimulus* in found in many areas of the world, from Chile to the Rocky Mountains (Wu et al., 2008). It is an appealing plant model due its relative ease of care and short generation time of two to three months. In addition to this, *Mimulus* has several fully sequenced genomes, as well as a stable transfection protocol using *Agrobacterium tumefaciens* (Ding and Yuan, 2016). Efficient transfections allow for gene characterization through plasmid transfections, RNAi-mediated gene knockdowns, and phenotypic rescue assays. Use of these genetic approaches are critical to holistically and conclusively analyze how a gene product impacts development. Ultimately, we seek to elucidate how a simple evolutionary framework for pattern formation allows minor changes in DNA sequence to impact the patterns created, resulting in the diversity found in nature.

A Novel Protocol for Phenotype Characterization

How can the interweaving genetics of development be studied in a large segregating population? To gain a cohesive understanding of pattern formation, we suggest the use of a phenotypic, genotypic, and mathematical approach to gene characterization. Such an approach would require a method by which phenotype can be scored in a consistent and objective manner using quantitative parameters as a first step. Phenotyping is the preliminary step of a

bulk-segregant analysis of an observable trait, such as flower patterns. Given that phenotypic studies often require large sample size to reduce noise in the data, this protocol must be modified to be high throughput. Quantitative data are used to separate the population into groups of similar phenotype.

Figure 4: Phenotype quantification workflow. Graphical representation our novel phenotyping protocol. Flowers are systematically imaged by dissection of top left, top right, and bottom center petals. After dissection, these petals are mounted on tape with an identification number. Pictures are taken with a Nikon D3200 equipped with a macro lens to ensure that details are detected. Images are resized to be of consistent dimensions using Adobe Photoshop. Here, the background is also removed so the petals are isolated in each image. Isolated petals are exported to ArcGIS (ESRI, Redland, CA USA), where each petal is converted to a binary map, or raster. These maps represent the localization of "red" and "non-red" on a petal. Binary rasters are imported into FragStats (McGargial et al. 2012) to calculate contagion and patch density. Binary rasters are also used to calculate proportion red with Python scripts, and are then used to calculate lacunarity. Lacunarity is calculated in ArcGIS using the Focal Statistics tool, a moving window average calculator that generates raw lacunarity. These raw data are used to calculate the final lacunarity with Python scripts. Once all metrics are collected, these data are used to separate a population by phenotype in a quantitative, objective manner.

The second step of the approach is to analyze the genes involved in creating these phenotypes. By using phenotype data to separate a sample population into groups, one can sequence the genomes of individuals in the groups to identify genes that may be responsible for the observed variation. The exact function of these genes can be characterized using RNA interference and gene knockouts, or use of CRISPR/Cas9 technology to see how changes in coding sequence and regulatory regions impact pattern formation.

The third step of this approach uses the identified patterning genes as a basis for mathematical modeling of pattern formation. Once the developmental genes have been characterized, mathematics can be used to predict how genetic changes may influence the

formation of patterns during development. These models have implications on developmental genetics, as one can model, for example, the impacts of increased activator binding on an organism's development, as well as evolutionary biology, using models to visualize how mutations influence traits such as pollinator preference. This holistic approach the analysis of pattern formation should be applicable for characterization of any petal patterns of interest, and perhaps useful for analyzing pattern formation in other organisms as well. Most importantly, however, this approach will elucidate the framework by which evolution generates such widespread diversity in the simplest manner possible.

Figure 5: Generation of a segregating F2 *M. luteus* population. *M. luteus* and *M. variegatus* are crossed to yield an F1 hybrid population (**top**). The flowers from the F1 population are then inbred through selfing to produce the divergent F2 population (**bottom**). Pictures courteously provided by Dr. Arielle Cooley of Whitman College.

This present study uses *Mimulus luteus* as a model for petal patterns, and we specifically focus on quantitative characterization of phenotypes in the population (Fig. 4). To do so, we use an F2 line of *M. luteus* that is segregating for petal design (Fig. 5) and characterize each plant's phenotype with a series of quantitative metrics. Experiments that deal with large populations and traits with high variation, such as patterning, often require large sample sizes; therefore, we present a novel, high-throughput protocol for quantifying phenotype in flowers using ArcGIS and FragStats. While these two softwares are typically used for landscape ecology, we co-opt these programs to create petal "maps" to characterize their patterns. ArcGIS is a geographic information software, and it is frequently used for mapping characteristics of large areas of land, such as a local ecosystem or the distribution of monetary aid in a developing country ("Who Uses ArcGIS?" 2009). Here, ArcGIS is instead used to create binary maps of each petal that describe where pigment is located with respect to the entire petal. Much of the process can be automated through features built into ArcGIS or the use of Python scripts, the language that the software runs on. Further data can be generated through FragStats, a landscape ecology software co-opted for petal trait quantification (McGarigal et al., 2012). FragStats uses image maps, or rasters, created in ArcGIS to calculate various metrics that describe the distribution or aggregation of pigment on a petal. To demonstrate this technique, we use a *Mimulus* population segregating for flower patterns and quantify the phenotypes using a variety of metrics.

Our results demonstrate the efficacy of our high throughput protocol. We find that, in our segregating *Mimulus luteus* population, there are several unique aspects of petal pattern formation that, to our knowledge, have not yet been characterized. First, we find that petals on the top of the flower appear to develop patterns similar to each other, but these patterns are

significantly different from patterns formed on the petals at the bottom of the flower. Second, we demonstrate that our population can be separated into groups by pattern morphology. Third, we find that genetics are a critical component to the types of patterns formed during development. We ultimately conclude that there is a strong genetic basis for pattern development, and that there are likely unique or interplaying genetic mechanisms for forming patterns on different petals within a flower. Our protocol is effective and low cost, allowing for pattern-based phenotyping of large populations efficiently. Future work will focus on the characterization of the genes responsible for pattern formation through gene sequencing and genetic analyses, as well as computational modeling of pattern formation during development.

Methods

Genetic Lines and Plant Growth

Crossing of *M. cupreus* by *M. variegatus* yields an F1 hybrid population. Generation of this F1 line is critical because it yields individuals that are heterozygous for alleles in both *M. cupreus* and *variegatus*. This F1 was further inbred through selfing; flowers from within the population were cross-pollinated, yielding an F2 line (Fig. 5). All crossing was done at Whitman College in the lab of Dr. Arielle Cooley, who graciously provided seeds of the F2 line. The F2 line shows dramatic segregation in flower phenotype, suggesting that recombination led to divergence of inherited alleles between plants. Plants are grown in a temperature-controlled growth room, where they are regularly irrigated until bloom. Sodium lamps supply light during the germination and growth phases.

Petal Imaging and Tissue Collection

Upon blooming, flowers are harvested from each plant. When a flower is harvested, an identification flag is used to identify that plant for future reference. For example, the plant that supplied the first flower is flagged as "Plant 1." We record flower sampling to ensure that each plant is sampled a sufficient number of times. Ideally, three flowers are taken from an individual. Sampling multiple flowers compensates for noise within a single plant, increasing the resolution of the phenotype analysis. If three flowers are not available at the time of harvest, then as many samples as available are taken, and a note is made to sample further in the future.

Pictures or petals are taken with a Nikon D3200 camera, using an AF-5 Micro NIKKOR 60mm lens. To control for lighting and color saturation, light is restricted to a single 60W bulb; all other lights are turned off to ensure only one light source is present during photography.

Flowers are dissected and mounted onto tape in a systematic manner (Fig. 6A). Colored tape provides contrast between the petals and the background, allowing easier petal isolation in Adobe Photoshop. Three petals, consisting of top left, top right, and bottom center (L, R, and C, respectively), are removed and mounted onto colored tape (Fig. 6B). On the side of the tape, an identification label is assigned to each flower, corresponding with P##F#, where P is plant number and F is flower number. Plant number corresponds to the identification flag in the donor plant, allowing for easy tracking of how many flowers have been taken from each plant. Flower number corresponds to the number of flowers taken from a given plant. Underneath the tape is an NGK color card, used to correct for color saturation across rounds of imaging. A ruler is placed above the petals and provides scaling control.

After a flower is imaged, a leaf sample is collected and frozen on liquid nitrogen; from there, it is stored on ice until it can be placed in a -80ºC freezer. This tissue sample will be used for DNA extraction in the future.

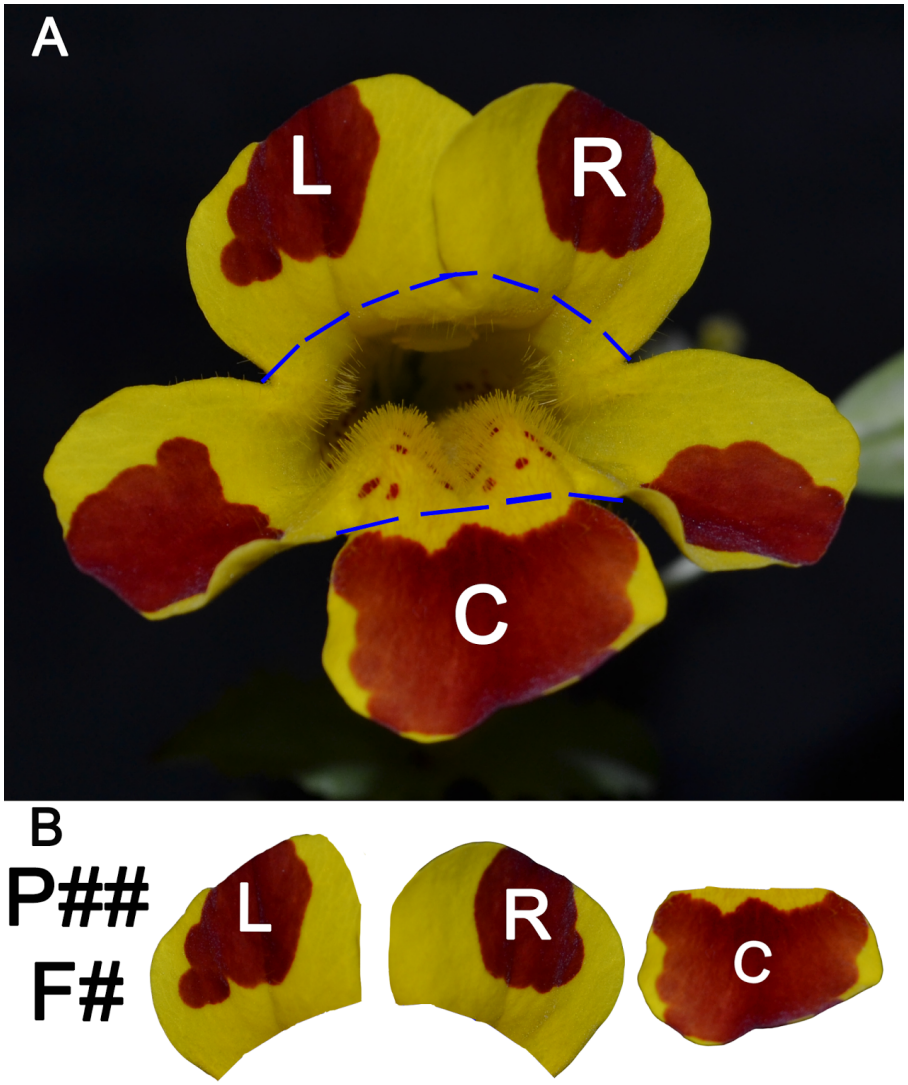

Figure 6: Flower petal dissection. **A:** The top two petals, and the bottom center petal, are dissected out along the blue dashed lines. Once removed, the petals are mounted on tape in the order specified in panel B. **B:** Plant number and flower number, used for identification purposes, are recorded on the left side of the tape.

Image Processing in Adobe Photoshop

Images are uploaded and named after their respective flower (e.g. p100f1.jpg). Before phenotypes can be quantified, flower petals must be isolated and converted into a binary format. Adobe Photoshop is used to rescale each image and remove the background from behind the petals (Fig. 7A). Here, rescaling of each image to a consistent size controls for variable distances between the camera and the stage; if this is not done, then patch size may be distorted due to overall irregularity of petal size between images. One control image, containing both the ruler and color card, is used as the size standard. The ruler and color card both serve as size constants, ensuring that all images are exactly the right size to keep proportions between petals correct. The opacity of the control image is reduced in Photoshop, and it is overlaid onto whichever image is being processed, allowing resizing to match dimensions of the ruler without disrupting the proportions. Once the image of interest matched the control, the control layer is deleted.

Once images are a uniform size, petals are isolated from each image; again, we use Photoshop for this step. Isolation of the petals from the background requires use of the Magnetic Lasso tool. This tool uses contrasting pixels to define the outlines of the selected target. The colored tape under the petals creates contrast to define the petal's border. Once all three petals are selected, the remaining background is deleted, leaving only the petals in the image (Fig. 7B). If any background remains that the lasso failed to detect, such as shadows near the edge of the petal, petals are further isolated with the eraser tool. Finally, the image is converted to a .tif file, the optimal file type for ArcGIS.

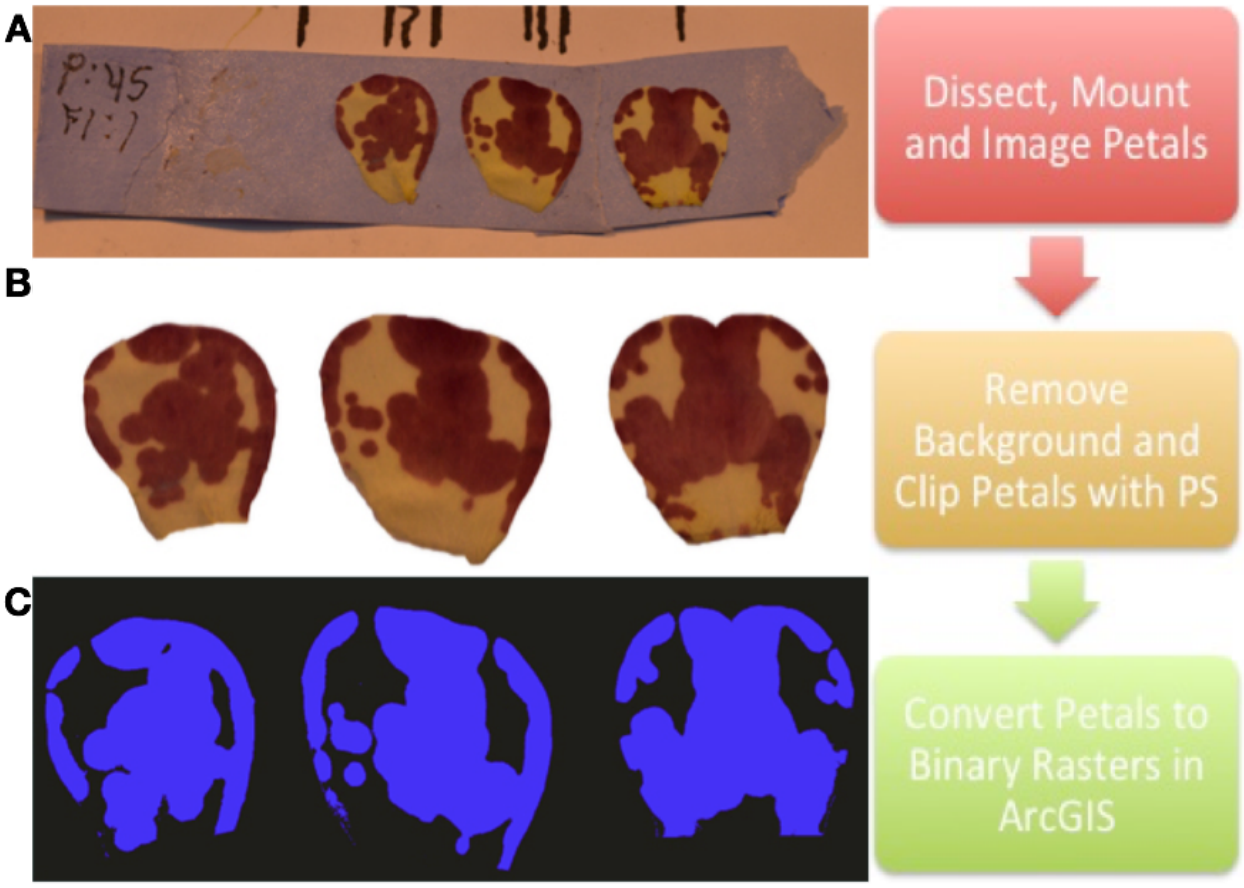

Figure 7: Petal images are processed using Adobe Photoshop and ArcGIS. **A:** Petals are dissected as outlined in Figure 6 and mounted onto tape. Once mounted, the petals are imaged. **B:** After imaging, petals are isolated from the background in Adobe Photoshop ("PS") and saved as .tif files for use in ArcGIS. **C:** In ArcGIS, images are converted into binary rasters, which map out the location of pigment in reference to the whole petal.

Raster Conversion in ArcGIS

Once petals are isolated and converted to .tifs, they are analyzed in ArcGIS. Here, petals are converted to binary rasters, which define pigment location through binary code and are used for all future calculations. In an unprocessed image, pixels are assigned a grey scale value ranging from 0 (absolute black) to 255 (absolute white). The Identify tool in ArcGIS reveals the value of any pixels of interest. This tool is used to designate a color threshold for binary raster calculation. Above this specified threshold, the computer converts non-red pixels to a value of 0;

below this threshold, a value of 1 is assigned to red spots. ArcGIS uses this converted input to output a binary raster file (Fig. 7C).

At this point, all three petals from a single flower are in the same image, and these petals need to be divided into their own individual images. To do so, each petal is clipped using a shapefile, a polygon rendered in ArcGIS that can be reshaped to fit any image. Modifying the shapefile to fit each petal ensures that only one petal is extracted, and no extraneous pixels separated from the petal are included in subsequent calculations. Once the shapefile overlaps exactly one petal, the petal is exacted from the rest of the image into its own layer using the Extract by Mask tool. When the petals are clipped into their respective image, each file is named P##F#X (where X defines L, R, or C) to identify the petal's location in the original flower.

Quantification of Phenotype in ArcGIS and FragStats

Clipped petals are then used for a multitude of analyses. Lacunarity, a measure of pattern randomness or "gapiness" in a local neighborhood (Plotnick et al., 1996), is calculated in ArcGIS using the Focal Statistics tool. This tool calculates pattern uniformity using a moving window average calculation, representing the overall cohesiveness of the pattern on each petal (Peterson and Hoef, 2010). The size or radius of the moving window is adjustable. Here, we run Focal Statistics seven times, each as a circular window with a different radius. Beginning with a small radius allows us to measure the small details of petal patterns. Increasing the window size represents "zooming out," and ensures that we measure the patterns from small details up to larger sections of the petal (Allain and Cloitre, 1991). Doing so creates a holistic understanding of the randomness and regularity of the pattern at various neighborhood size.

Any tool in ArcGIS may be automated using the Model-building tool, an application that lets the user map out automatic processes and designate an output file name and location. Processes can be planned out and automated through the model feature, and it does not require any prior coding knowledge. Here, lacunarity calculation is automated by designating seven different output rasters per petal, each raster representative of a different "neighborhood size." Neighborhood size corresponds to the size of the moving window's radius, and varying the size creates a narrow to holistic quantification of pattern randomness. The lacunarity output rasters are no longer binary. Instead, the pixel values calculated by Focal Statistics form localized gradients that represent the proximity of pigment in local areas and the regularity in the patterns.

Lacunarity rasters are converted into ASCII files, a table text file. In these files, each pixel is converted into its numerical value calculated by Focal Statistics. To measure total lacunarity, we used Python codes that automatically averaged the lacunarity in each petal. By automating both the Focal Statistics and lacunarity collection steps, we created a high throughput pipeline for lacunarity measurement in each petal. These data are used to separate flowers, and plants, into groups by pattern morphology in preparation for the bulk segregant analysis.

The clipped binary rasters created by the Extract by Mask tool are also useful for phenotyping. Conversion of these files to ASCII format allows importation into FragStats, a landscape ecology software that we co-opt to calculate metrics of pigment patching on our petals. While normally used to measure landscape and edge patches in maps, we instead use FragStats to measure aggregation and dispersal of pigment across each petal. We calculate two metrics: patch density (PD) and contagion. PD compares the number of patches to the squared area of the petal, representing the general dispersal of pigment. Since PD is calculated based on

the number of patches, a large single spot would result in a small PD; little, plentiful patches result in a higher patch density. Therefore, we use this metric to represent the spottiness of each petal. Contagion is an aggregation metric that measures how patches are packed in the center of the petal. It is important to note that this is not the inverse to PD, but rather a novel metric representing the localization of the patches. FragStats automatically analyzes each image in its queue and outputs the data into a table for each petal. As such, this process is automated and relatively quick depending on the number of samples run at a time.

Clipped rasters are used for one more round of data collection: total petal area and proportion red. Since rasters are in a binary state, a simple Python script can calculate these metrics. Total petal area sums the number of pixels in a raster. Proportion red sums the number of red pixels ("1's") and divides this by the total petal area. Total area is not used for phenotyping, but must be calculated to determine the proportion red. Proportion red, while a simple metric, is important for measuring overall pigment content in a petal.

In total, we collect data on lacunarity at seven window sizes, contagion, patch density, and proportion red. Using these metrics, we can quantitatively analyze the variation within and between plants using simple statistical analyses.

Data Analysis

Data were analyzed using several software platforms. SPSS Statistics is used to calculate linear mixed model results. JMP, software developed by SAS Institute, is used for linear regression analyses, ANOVA, T-tests, and principal component analyses. Figures are created using Adobe Photoshop and Graphic.

DNA Purification

DNA is extracted from the tissue collected as previously mentioned. Until extraction, the tissue is stored at -80ºC in a microfuge tube labeled with the donor plant's ID number. Extraction is done using two different methods: the detergent-chloroform protocol Plant DNAzol, and the Qiagen Plant DNeasy column kit. DNAzol was used initially due to high yield with larger tissue samples, but the switch to Qiagen improves yield with smaller tissue samples and overall purity.

Yield is quantified through Qubit fluorometry, Nanodrop analysis, and gel electrophoresis. Fluorometry measures fluorescence emitted by a DNA-bound dye after laser excitation and determines DNA concentration by the level of fluorescence. Nanodrop analysis can also determine concentration, but is known to overestimate actual concentration (Simbolo et al., 2013). Instead, Nanodrop is used to determine DNA purity, such as the amount of proteins, RNA, and free salts in solution with the DNA that may disrupt sequencing. Electrophoresis measures DNA integrity. During extraction, DNA can be sheared through frequent pipetting or repeated freeze-thaws. To ensure that DNA was high quality, periodic gels were run with ethidium bromide staining as a quality control measure. Presence of a large band at a high base pair size in the gel represents high quality DNA; small bands or smears are indicative of either DNA fragmentation or RNA contamination. If any of these tests show low concentration or purity, then the Zymo Clean Up & Concentrator Kit is used to purify samples and resuspend them at a higher concentration. It is important to note that during Qiagen and Zymo purification, the final sample cannot be eluted in the kits' buffers. These buffers contain EDTA, and while this compound stabilizes DNA for long periods, it also interferes with restriction enzyme activity during DNA library preparation. All samples are suspended in deionized, sterile water.

Library Preparation for Genotyping-by-Sequencing

After quantification and purity assessment, pure samples of sufficient concentration are converted into DNA libraries. This is the final step before each sample can be sequenced, and it is essential for next-generation sequencing. During library preparation, samples are digested by the restriction enzyme ApeKI, which produces overhanging sticky ends at DNA cut sites randomly across the genome. Adapters are ligated to these sticky ends. Adapters, provided by the Twyford Lab, are ssDNA oligonucleotides with one end complementary to the sticky ends. The oligos base pair to the fragmented DNA and are ligated by T4 ligase. Once ligated, these adapters bind our DNA to flow cells during sequencing. Each adapter also contains a short unique sequence that will correspond to a sample after sequencing. During adapter ligation, each sample is assigned a specific well of adapters, with each well containing a novel barcode sequence. These barcodes are compatible with use of Restriction-site Associated DNA (RAD) marker sequencing. RAD-seq reduces sequencing costs and is a high throughput approach to sequencing large populations in a single sequencing reaction. When sequenced, these barcodes can be cross-referenced back to the donor plant's identification tag, telling us "this sequence is from Plant X."

Once T4 ligase covalently bonds the adapters, the adapter-DNA fragments are elongated and amplified with a truncated PCR protocol. To amplify adapter-ligated fragments, we use primers complementary to a sequence on the adapters. The reduced cycles prevents introduction polymerase-induced errors and subsequent amplification of these errors. Finally, the amplified libraries are purified during a size-selective AMPure bead clean-up protocol. Bioanalyzer chip analysis confirms presence of libraries and distributions of various fragment sizes.

Results

Distinct Variation between Top and Bottom Petals

The goal of this present study is to provide evidence that pattern formation during flower development follows a genetic framework, and that mutations in this framework may be the basis for rapid evolution of a myriad of flower patterns. Using our high throughput phenotyping protocol, we ask three questions. First, how do patterns vary between petals? Second, is genotype a significant component contributing to the observed pattern variation in a divergent F2 population of *Mimulus luteus*? Third, can our protocol separate a population into groups by morphology? Here, we present data collected by phenotyping of 146 flowers from a total of 83 plants.

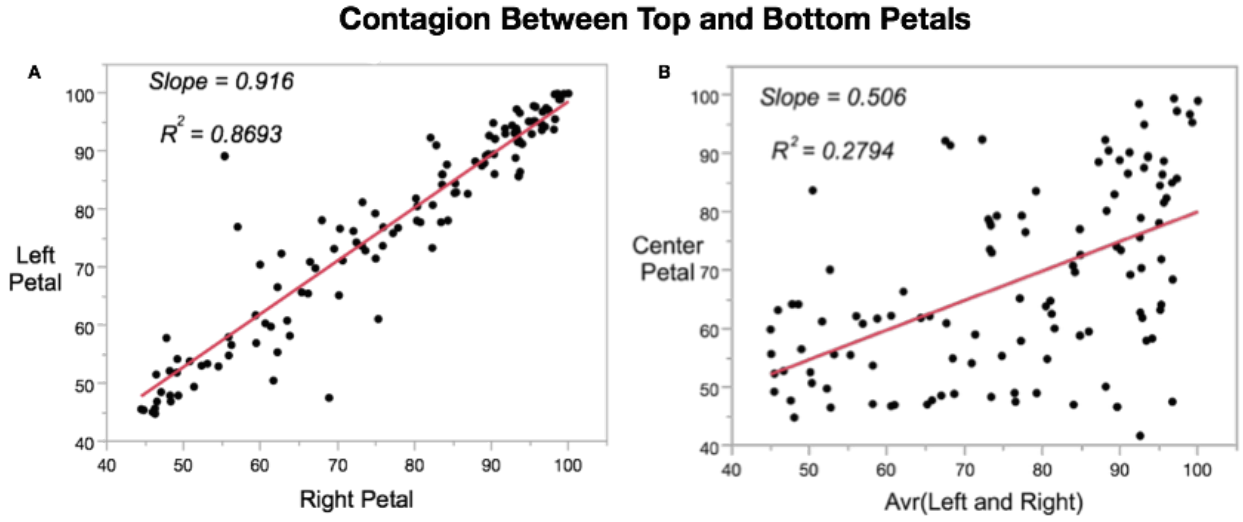

Figure 8: Contagion is highly correlated between the top two petals, but is not related to bottom petals. Contagion generally translates to central aggregation of pigment. **A**: Linear regression of left petal by right petal shows that contagion is highly correlated between top petals ($r^2 = 0.8693$). **B**: Contagion is not as highly correlated between the bottom center petal and the average contagion of the top two petals ($r^2 = 0.2794$).

Contagion in Top Petals is Distinct from the Bottom Petal

Contagion represents the amount of pigment localized in the center of the petal. In FragStats, this is defined as "the inverse of edge density". Specifically, it is known as an interspersion metric, referring to the "intermixing of patches of different types, and [this metric] is based solely on patch adjacencies" (McGarigal et al., 2012). In more complex analyses, such as those done in landscape ecology, contagion can measure the relative proportion of multiple patch types, analyzing complex landscapes such as the interspersion of lakes in a forest. However, we are only interested in two patch types, red and non-red, so FragStats runs the binary raster ASCIIs for only two types of patches. For our purposes, contagion illustrates how centrally aggregated pigment is on a given petal. A lower value for contagion represents more centrally localized pigment, while high contagion represents edge-localized pigment. Simply, a low contagion refers to a high aggregation of pigment in spots where pigment is found. It does not represent overall pigment density, nor the size of the patches, but merely the manner in which patches aggregate, and their relative position to the edge of the petal.

To determine if pigment aggregates similarly between the three petal types (left, right, and center), we measure contagion in each petal and tested the correlation between the petal types (Fig. 8). We find that there is a strong positive correlation between left and right petals for contagion ($r^2 = 0.8693$), indicating that the top two petals develop similarly aggregated patterns during flower development (Fig. 8A). In contrast, the center petal shows notable dissimilarity between the top two petals when the mean contagion of the top two petals is compared to the bottom center petal (Fig. 8B). Here, the center petal is not nearly as related to the top two petals

(r^2 = 0.2794), indicating that the center petal creates different aggregation patterns than the top petals.

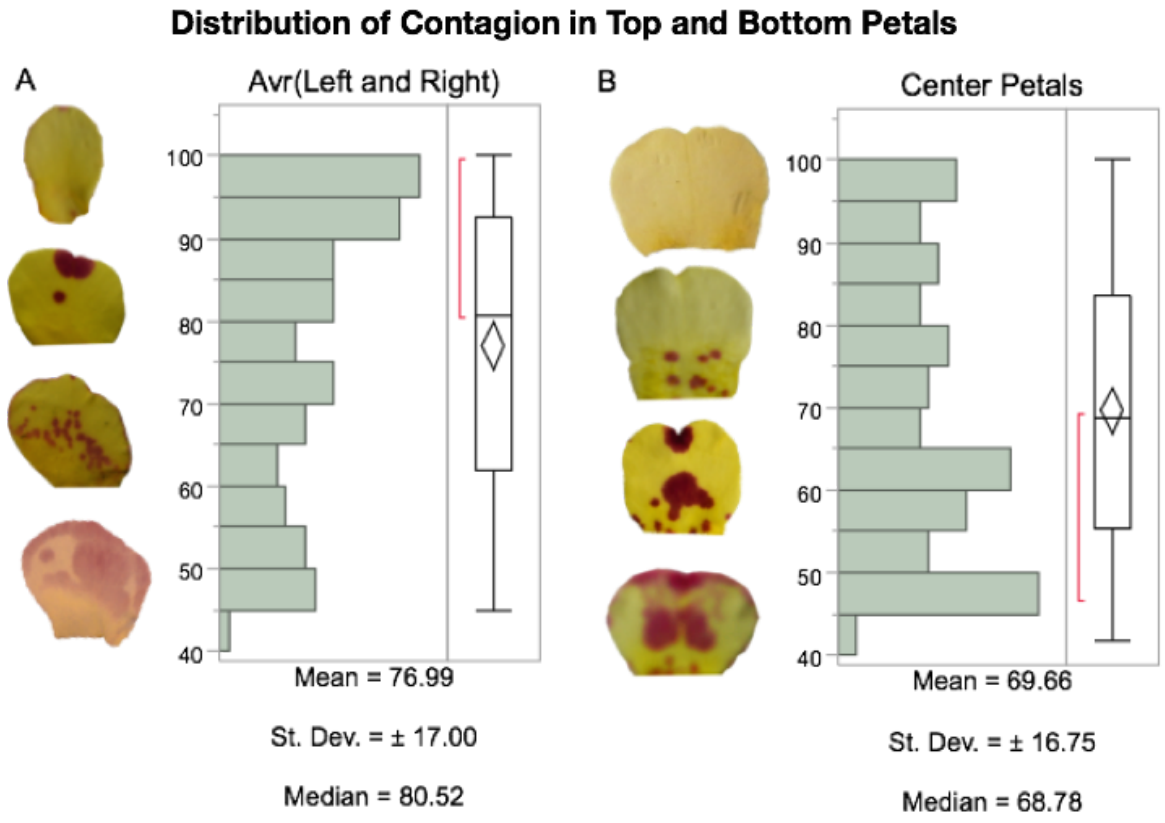

Figure 9: Broad distribution of contagion between all flowers. Contagion measures the central aggregation of pigment in a flower petal. High contagion indicates that there is little pigment localized in the center of the petal. Contagion is broadly distributed in both top (**A**) and center (**B**) petals. Overall, center petals have more centrally aggregated pigment than top petals.

We calculate the distribution of contagion between the top and bottom petals as well. Given the high correlation between left and right top petals, we measure the distribution of the mean top petal contagion and the distribution for center petals (Fig. 9). Petals corresponding to different contagion values are plotted alongside their position on the distribution. The mean contagion for the top petals is higher than that of the bottom center petal; therefore, top petals have less central pigment aggregation than the bottom petal. This is clear when looking at the

flowers, as bottom petals tend to have a larger proportion of pigment in the center. Top petals have more points at a high contagion value, with 50% of the data falling within the range of 81-100 (Fig. 9A, red bracket). Center petals have a more widespread distribution, with 50% of contagion falling within the range of 46-69 (Fig. 9B, red bracket).

These data suggest that top petals create patterns that are distinct from the bottom petal, as top petals aggregate their pigment towards the edge of the petals, while center petals localize their pigment in the middle. There may be two independent systems for localizing pigment between top and bottom petals during pattern formation. However, we cannot determine this without genomic data.

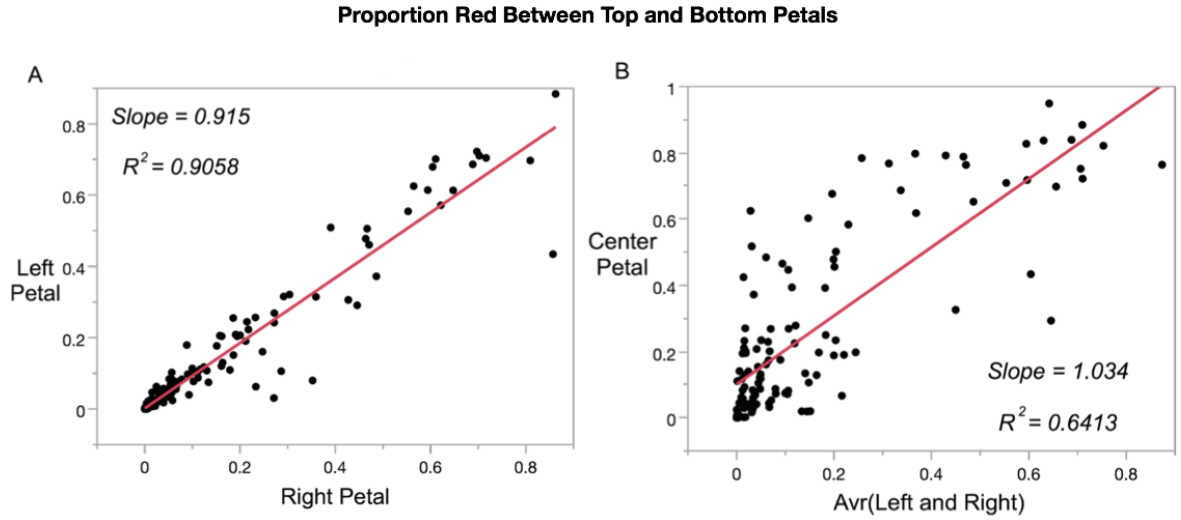

Figure 10: Proportion red is highly correlated between top two petals, and bottom is related to top two petals, but to a lesser specificity than in top petals. Proportion red is the ratio of area red:total area. **A**: Proportion red is highly correlated between the left and right top petals ($r^2 = 0.9058$). **B**: Proportion red of the bottom center petal shows some relationship to the average proportion red of the top two petals ($r^2 = 0.6413$); however, this relationship is not as strong as that between the top petals.

Proportion Red in the Top Petals is Distinct from the Bottom Petal

$$PR = \frac{red\ area}{total\ area}$$

Proportion red is a simple metric, representing the percentage of petal colored red compared to the total surface area. This provides insight into the amount of anthocyanin produced within a plant, which may be useful for dividing plants by phenotype. It is also important as supplementary information for other metrics, such as lacunarity, which do not discern between pigment colors during calculations. Proportion red does not provide data regarding where the pigment is localized or how densely it is dispersed.

We analyze proportion red between petals using a linear regression (Fig. 10). Proportion red is highly correlated between the left and right petals ($r^2 = 0.9058$). This suggests that the top two petals may control anthocyanin biosynthesis by the same mechanism to create similar patterns on each petal. The bottom petal proportion red is somewhat related to the top petals ($r^2 = 0.6413$). There is a positive relationship between petal types, but this model demonstrates that there is some difference between the amount of anthocyanin produced between the top and bottom petals.

Proportion red is less widely distributed compared to contagion, with most petals having <50% red (Fig. 11). Center petals have a higher average proportion red; on average, 25% of the petal surface expresses anthocyanin. This mean is higher than the mean the top petals, which average 15.8% red. Despite this, regardless of petal location, the majority of petals are less than 20% red (Fig. 11, red brackets).

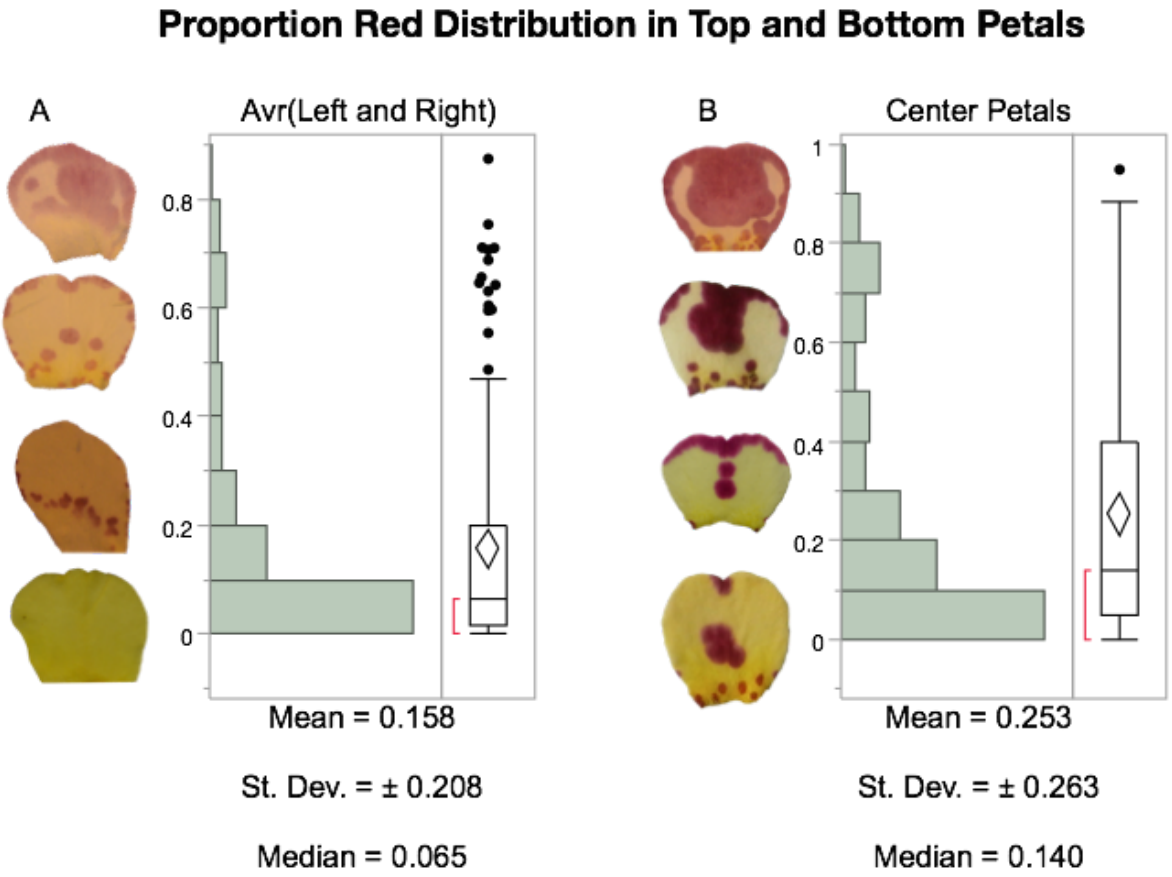

Figure 11: Distribution of proportion red is similar between top and bottom petals. Proportion red is calculated by the area of red patches divided by the total area of the petal. Both top and bottom petals tend to be below 20% red. Bottom petals express more anthocyanin than top petals on average.

Overall, the bottom petal displays more red pigment than the top two, with a much wider range of pigment expression in the bottom petals. Therefore, there may different systems controlling pigment biosynthesis between the top and bottom petals.

Patch Density is Similar Across All Petals

$$PD = \frac{number\ of\ patches}{total\ area^2}$$

Patch density refers to how frequently patches are placed within a unit area. The density is calculated by measuring the ratio of the number of patches to the squared total area of the

petal. This differs from contagion, the interspersion metric, by representing the average density of pigment in a picture, as opposed to the pigment's localization. As such, it does not take spot size into account. We use this metric as a representation of petal "spottiness."

How does patch density relate to these previously described metrics? Contagion measures the localization of the patches, but not the density of them. Proportion red only identifies amount of pigment, but does not provide any information regarding localization or dispersal. Patch density measures the spread of pigment, but not necessarily where in particular patches are in relation to each other. Density allows us to measure how the pigment disperses, while contagion tells us how pigment aggregates. It also loosely translates to size of patches when compared to proportion red, because if petal has a high proportion red but a low patch density, we can conclude that the pigment is localized to only a few large spots.

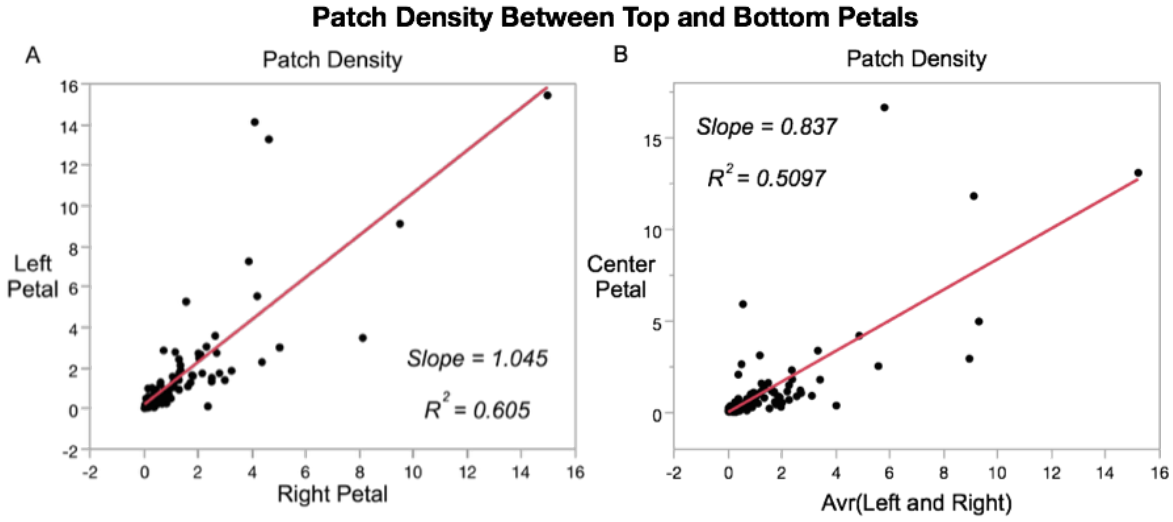

Figure 12: Patch density is correlated across all petals. Patch density measures the number of patches over the squared total area. This represents the density of patches (spottiness) regardless of patch size. **A**: Patch density between left and right petals shows a moderate positive correlation ($r^2 = 0.605$). **B**: Patch density is similarly correlated between the mean PD of top petals and the bottom center petal, but to a slightly lesser degree than left-to-right PD ($r^2 = 0.5097$).

Patch density is not very different between petal locations (Fig, 12). There is a strong correlation between the left and right petal ($r^2 = 0.605$). Much of the data falls in the lower range, indicating that patches are not very dense in most top petals. However, there is also a strong correlation between top and bottom petals as well ($r^2 = 0.5097$). While this correlation is not as strong as the correlation between the top petals, there appears to be a similar mechanism between the dispersion of patches between the top and bottom petals. Because most data falls into the lower ranges of patch density, we did not generate distributions of patch density for either petal type. The fact that there is no distinct difference between top and bottom petals suggests that the same locus controls patch density across all petals during development, or that the top and bottom petals have no distinct mechanism through which they generate patch density.

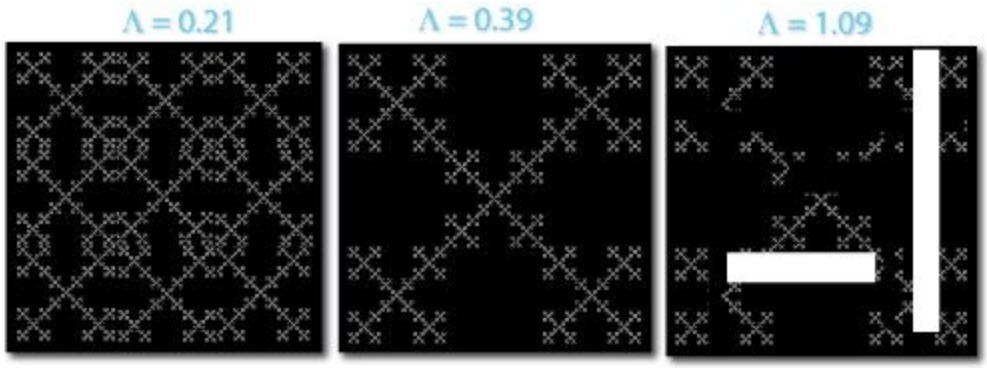

Figure 13: Lacunarity measures fractalized pattern regularity. Lacunarity calculates randomness in a sample through moving window averages at various window sizes, or radii. Low values represent dispersed patterns of color blotches regardless of the window size used (left panel). As the pattern of blotches aggregates, lacunarity increases (middle and right panels). Open source image from Wikimedia Commons courtesy of Audrey Karperien.

Localized Lacunarity is Distinct Between Top and Bottom Petals

Lacunarity measures the randomness of a pattern based on the fractalization of the patterns (Fig. 13). Non-changing lacunarity values represent fractal patterns that do not change

as one zooms in or out on the pattern. As the patterns become more aggregated, lacunarity values rise. Regularity at different "zooms," or neighborhood sizes, is determined by moving window averages. A window is a predefined area of pixels, and this window moves across the image and determines the sum of red and non-red pixels within its radius. As the window moves across the image, this sum changes based on the regularity of the pattern. Changing the radius of the window increases the neighborhood size, representing a change in zoom. Small radius windows examine the details of a pattern, while large radius windows look at the overall pattern across the image. Lacunarity values can be compared between window sizes to represent how regular the pattern is at any given zoom.

When calculating lacunarity, we measured lacunarity at seven different zoom, or window radii, which produces seven different values for lacunarity. We designated a circular window with a radius of 1, 5, 10, 20, 40, 60, 80, or 100 pixels to calculate pattern regularity. Radius 1 is limited in value; small fluctuations in a pattern could result in drastic changes in lacunarity. At higher neighborhood sizes, the data is far more robust. Across the different neighborhood sizes (excluding r=1), most lacunarity values are highly correlated within a flower, so radius = 5 and 100 are used most for models, as these values represent the two extremes of pattern regularity (Fig. 14).

We created linear regression models for a window of radius = 5 to determine if there is a difference in pattern regularity between petals. Lacunarity is highly correlated between the left and right petals ($r^2 = 0.7557$, Fig. 14A). There is some discrepancy between left and right petals, but patterns are largely regular between the top two petals. At a window radius of 5, bottom petals are distinctly different from the top petals ($r^2 = 0.3537$, Fig. 14B) These data suggest that a

very dissimilar pattern is generated between top and bottom petals, while the left and right top petals are more similar in patterning.

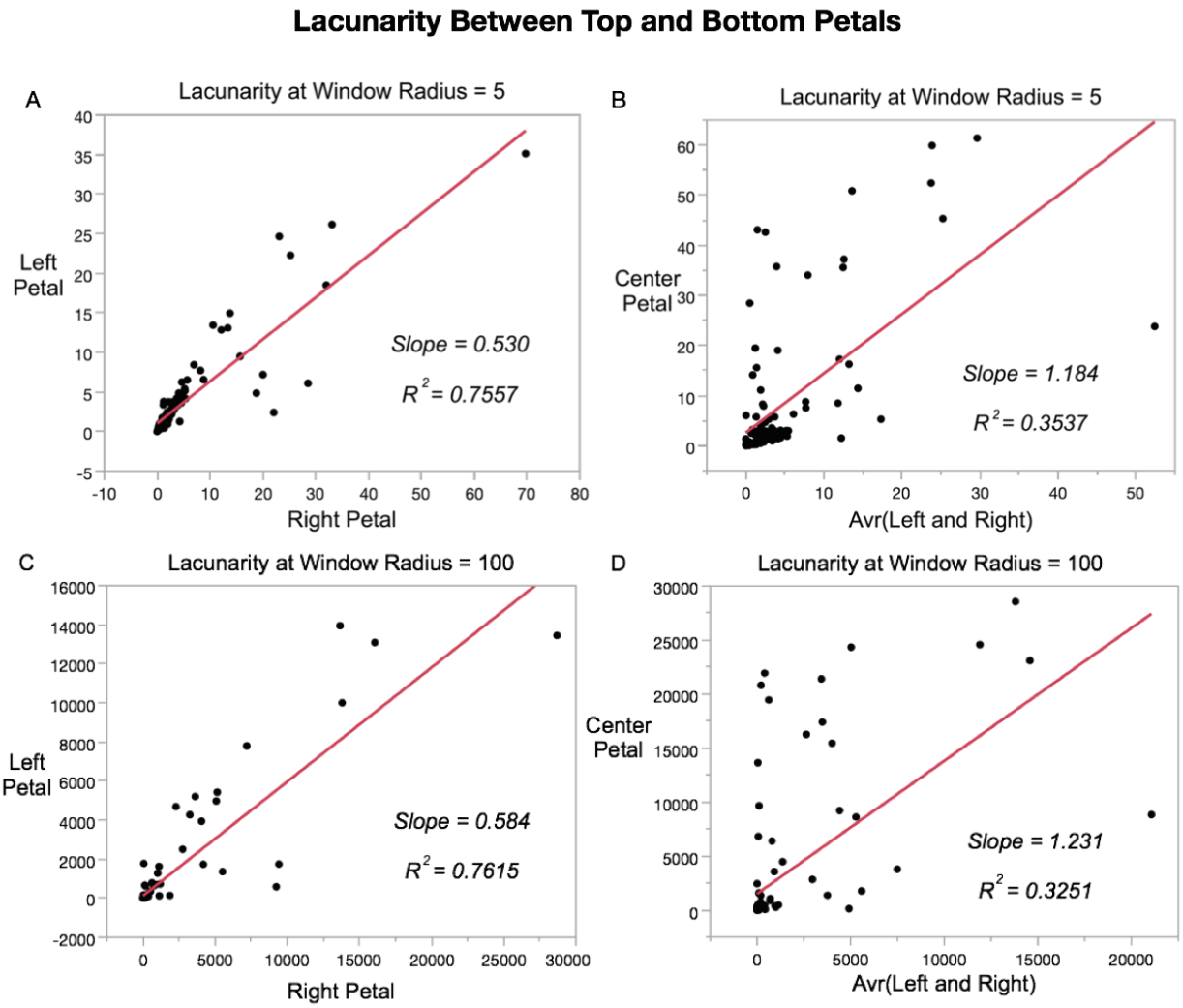

Figure 14: Lacunarity at small and large neighborhood sizes is correlated between the top two petals, but distinct between the top and bottom petals. Lacunarity is a measure of randomness or regularity within a pattern. Low lacunarity represents consistent patterning. Lacunarity is calculated by a moving window average. Changing the radius of this window represents a change in "neighborhood size," i.e. zooming out to observe a larger portion of pattern at once. **A**: Lacunarity of a small neighborhood (radius = 5) is highly correlated between the top two petals ($r^2 = 0.7557$). **B**: Lacunarity of a small neighborhood is weakly correlated between the top and bottom petals ($r^2 = 0.3537$). **C**: Lacunarity of a large neighborhood (radius = 100) is highly correlated between the top two petals ($r^2 = 0.7615$). This correlation is equivalent to the correlation seen in the small neighborhood analysis. **D**: Lacunarity of a large neighborhood is weakly correlated between the top and bottom petals ($r^2 = 0.3251$). This shows that pattern randomness is distinct between top and bottom petals regardless of neighborhood size analyzed.

To ensure that we understand the whole picture, we also investigate if a relationship exists between petals at a window size of radius 100. The increase in radius represents a "zoomed-out" measure of lacunarity, looking at larger portions of the petal at once. Again, left and right petals show a high correlation to pattern regularity ($r^2 = 0.7615$), while there is less of a relationship between top and bottom petals ($r^2 = 0.3251$; Fig. 14C and D). These r^2 values are very distinct between the two extreme window sizes, so we conclude that top and bottom petals are robustly different in terms of pattern regularity and overall patterning. It is interesting to note that twin comparisons at either window size (i.e. left vs. right at r = 5 or = 100) are almost identical in slope and r^2. Therefore, we conclude that data from lacunarity is robust at multiple window sizes, but we recommend that several lacunarity values be used when phenotyping to ensure that all aspects of pattern formation are considered.

Figure 15 illustrates how pigment patterns impact lacunarity values. At a small window size (r = 5), the logarithmic distribution is normal within the population in top and bottom petals (Fig. 15A and B). One facet of lacunarity to note is that this metric determines pattern regularity, but it is unable to discern which pigment type, red or non-red, is contributing the most to patterning. This is most obvious in Fig. 16B, where the petal with the lowest lacunarity has the highest proportion red of the petals on display. This fact demonstrates that lacunarity is not a stand alone metric for phenotyping, and that other metrics such as proportion red are required to parse out the pattern diversity found in a segregating population. At larger window sizes (Fig. 15C and D), there are broad lacunarity distributions in top and bottom petals. These data reaffirm our thoughts that using several lacunarity window sizes is useful for robust phenotyping, if used in conjunction with other metrics.

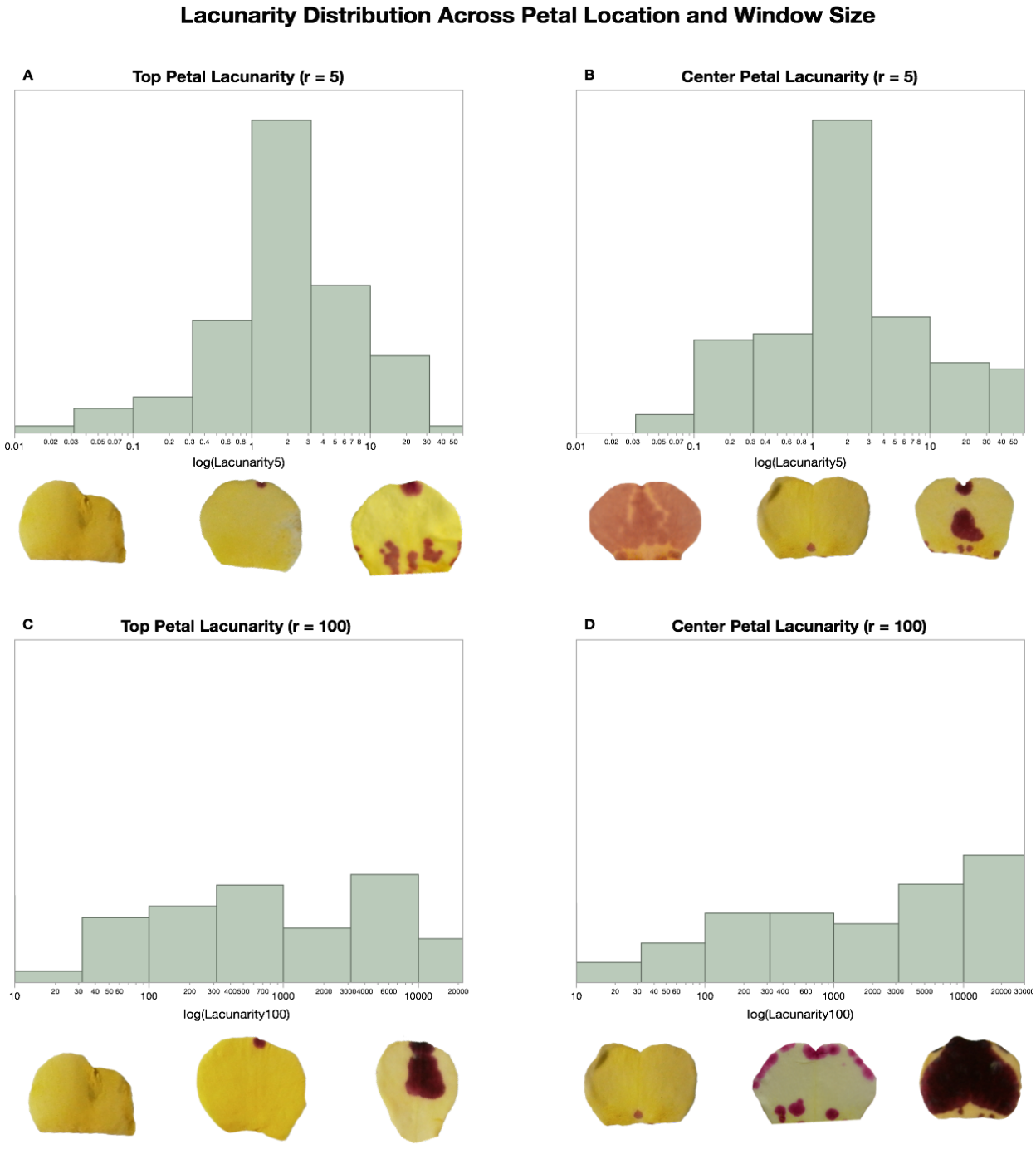

Figure 15: Lacunarity distributions are normal at small window sizes but broader at large window sizes. **A** and **B**: Logarithmic distribution of lacunarity for top and bottom petals at radius = 5. Note how the petal on the left in Panel B shows low lacunarity (high regularity) despite a high proportion red. This demonstrates that lacunarity does not factor in which patches contribute most to the pattern, as a homogeneous distribution of anthocyanin returns a low lacunarity value. **C** and **D**: Logarithmic distribution of lacunarity for top and bottom petals at radius = 100. This demonstrates that the distribution may change depending on window size, but some petals still return proportionally the same lacunarity. These data illustrate that lacunarity is effective at measuring pattern formation but is incapable of discerning between types of pigments. Therefore, phenotyping cannot be done with lacunarity alone, but requires several metrics to parse out differences in phenotype.

Top and Bottom Petals are Phenotypically Distinct

These data demonstrate that the top petals display largely similar patterns between the left and right petals, and these patterns are notably different from those found on the bottom center petal. While these correlations suggest that there is a difference in the formation of these patterns, we ask, are these observed differences in top versus bottom petals significant? A principal component analysis (PCA) for all flowers across all described metrics demonstrates that, in top and bottom petals, the first principal component contributes to a large portion of the observed variance at 51.6% (Fig 16A). These data suggest that a single factor, principal component 1, contributes to much of the variance observed in the population. However, it does not directly identify the source of this variation. Further analysis of variance (ANOVA) using T-tests comparing the PC1 values by petal location demonstrates that there is a significant difference in the contributions of PC1 between the top and bottom petals (Fig. 16B; Two-tailed T-test, $P < 0.0001$). We conclude that patterns formed on the top and bottom petals are significantly different, but this test does not provide insight into *how* the patterns are different.

To decipher these differences, we measure the variance in each metric to see which ones are contributing the most to this difference between petals. Again, top petal values were averaged to represent "Top," and bottom petal values represent "Center." Two-tailed T-tests analyze differences between the means of top and bottom petals at a confidence interval of 0.95. We used two-tailed T-tests because we expect there to be a difference in the means of these metrics (based on the PC1 results and linear regressions), but we cannot predict the directionality of these differences. Results are outlined in Table 1, and illustrated in Figure 17.

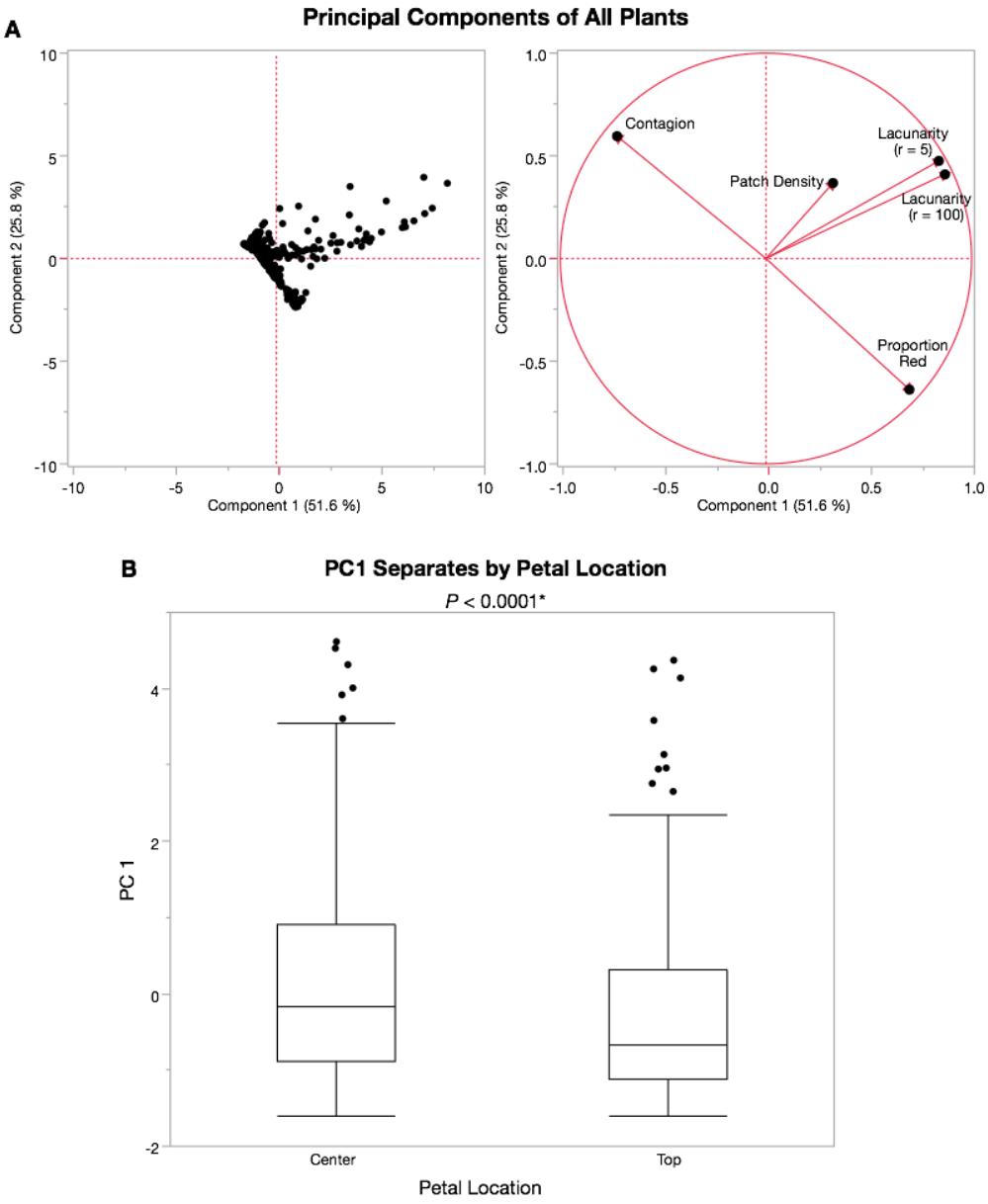

Figure 16: Top and bottom petal patterns are significantly different. **A:** Principal Component 1 explains 51.6% of the observed variance in the sample population across all metrics. **B:** Petal location (top vs. bottom) contributes significantly to PC1 (Two-tailed T-test, $P < 0.0001$*). These data are a strong indicator that different mechanisms control top and bottom petal pattern formation. At the least, patterns are significantly distinct between petal types.

	μ Top	SD Top	μ Center	SD Center	*P*-value
Contagion	77.1	1.07	69.4	1.48	<0.0001*
Prop. Red	0.161	0.0141	0.258	0.0197	0.0002*
Lacunarity (r = 5)	4.08	7.22	6.97	12.78	0.0147*
Lacunarity (r = 100)	989.25	265.94	2530.72	371.37	0.0061*
Patch Density	1.229	0.135	1.046	0.187	0.419

Table 1: Top petal patterns are significantly different from bottom petals. Two-tailed T-test analysis results for each metric by top and bottom petals. * indicates significance at a confidence interval of 0.95. In all metrics except patch density, the top petals are significantly distinct from the bottom center petal. Within group variation does not explain the patterns observed here. These data suggest that there may be independent or entirely separate regulators for the formation of patterns in a petal-dependent manner, based on the fact that patterns between petals are so distinct.

Means from three of the four metrics used are significantly different between top and bottom petals (Fig. 17). While this is hinted at in the correlation analysis, the T-tests provide clear evidence that the top and bottom petals create unique patterns during development. Contagion and proportion red are most distinct (Two-tailed T-test, $P < 0.0001$ and 0.0002, respectively; Fig. 17A and B). Lacunarity is split into two measures, for window radius 5 (Fig. 17D) and 100 (not shown); however, at either window size, the top petals show distinct patterning compared to the bottom petal (T-tests; $P_5 = 0.0147$*; $P_{100} = 0.0061$*). Patch density is the only parameter that is not significantly different between groups (T-test; $P = 0.419$; Fig. 17C). We conclude that patterns vary between top and bottom petals, but not between the top left and top right petals. Specifically, the top petal patterns are less centrally aggregated (higher contagion), have less pigment, and are more regularly formed (lower lacunarity) when compared

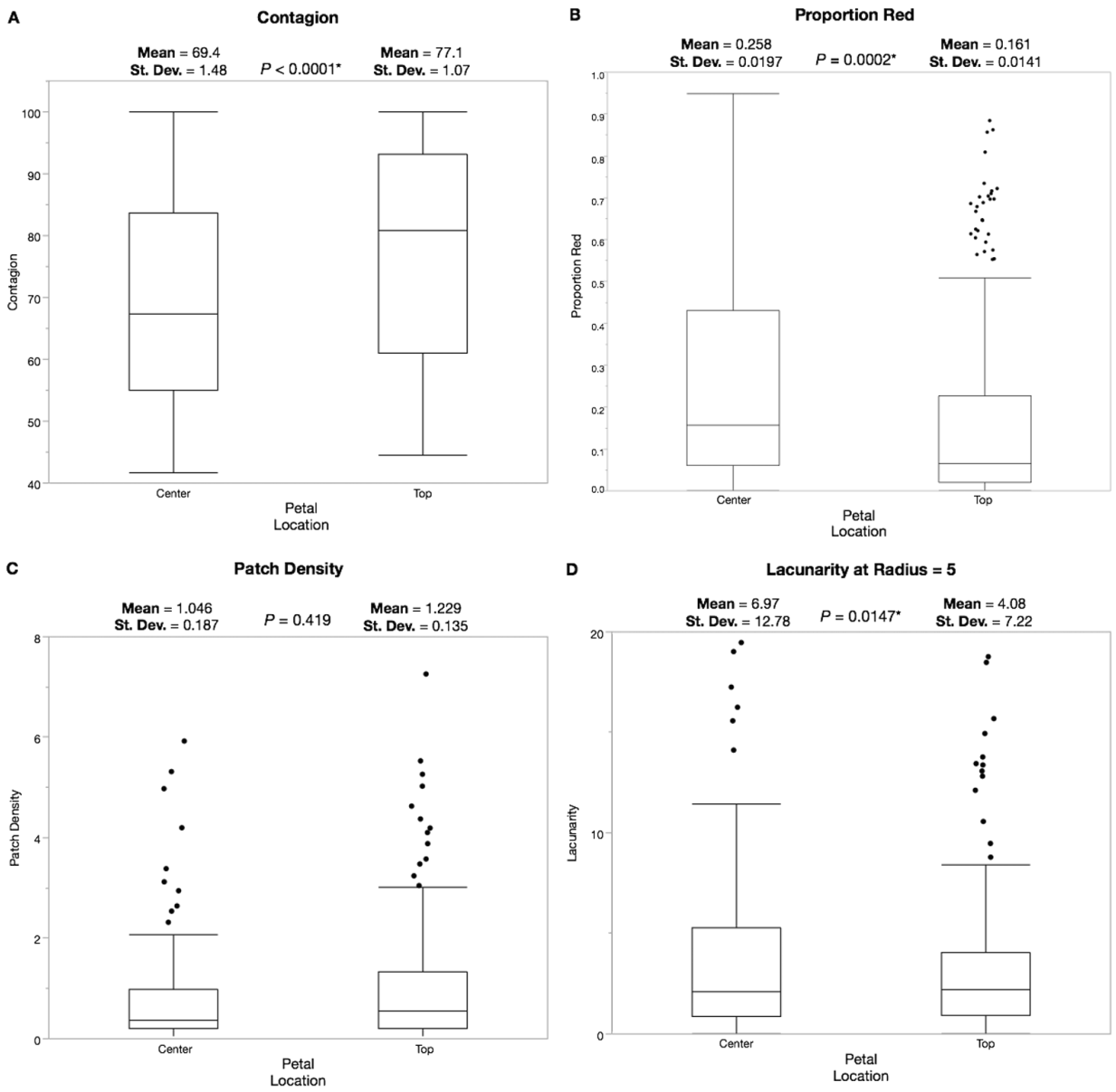

Figure 17: Patterns between top and bottom petals are phenotypically distinct with regards to central pigment aggregation, total amount of anthocyanin, and uniformity of the pattern. T-test results between top (left and right) and bottom petals for contagion, proportion red, patch density, and lacunarity at a window size of 5. **A:** Top petals (n = 260) have a higher mean contagion, translating to reduced pigmentation in the center of the petal. The bottom center petal (n = 136) has more pigment centrally localized, as seen by the significantly lower contagion (Two-tailed T-test; $P < 0.0001$*). **B:** Center petals (n = 141) have a higher mean proportion red than the top two petals (n = 273; Two-tailed T-test, $P = 0.0002$*). **C:** Patch density does not differ significantly between top (n = 260) and bottom (n = 136) petals (Two-tailed T-test; $P = 0.419$). **D:** Mean lacunarity at a small neighborhood size differs significantly between top (n = 273) and bottom (n = 140) petals (Two-tailed T-test; $P = 0.0147$*). Top petals have a lower mean lacunarity, suggesting more uniformity in their pattern formation. The bottom petal has a higher mean lacunarity, representing more disrupted patterns at even the closest "zoom" of the patterns. **Not shown**: Lacunarity at r = 100 followed the trend seen in window size = 5 between top and bottom petals (Two-tailed T-test; $P = 0.0061$*).

to the bottom petal patterns. This suggests that there are different genetic mechanisms behind pattern generation in the top and bottom petals.

Variation in Pattern Formation has a Strong Genetic Component

Now that we have established that patterns on the top petals are distinct from the bottom petal patterns, we ask the question, how does genotype contribute to pattern formation? We investigate this by partitioning petals into groups in the same manner as with the T-tests: top and center. With these groups, we create a linear mixed model to determine what variables contribute the most to the observed phenotypes.

A linear mixed model is a multivariate analysis that determines proportional contributions of specified variables to phenotype. One benefit of these models is the freedom to use interdependent variables. In a standard linear regression model, all variables must be measured independently, meaning that no samples can overlap in the analysis. In a mixed effect model, interdependent samples, such as multiple metrics from the same flower, are assigned an identity as either "fixed" or "random." The equation accounts for the minor differences between individual interdependent samples, and considers these small discrepancies when calculating error between samples.

Each metric we have described thus far (contagion, proportion red, patch density, and the two lacunarity measures) is modeled as the following function:

$$Metric = Plant + Flower + Petal(top\ v.\ bottom) + \varepsilon$$

Metric is a parameter we measured, such as contagion, and is a dependent variable to compare variance described by the following variables. *Plant* refers to the Plant ID. In this model, we treat *plant* as a "random" independent variable; we cannot conclude that it is fixed

because we have found no evidence thus far that metrics correlate across plants. *Flower* is treated as "random" and independent as well, because we do not expect to see observable differences between flowers within the same plant. *Petal*, on the other hand, identifies as a "fixed" independent variable due to the clear differentiation of top versus bottom petal determined by the previously described analyses. Finally, ε represents control for general error that is undefined, such as sampling error or random drift. Results are summarized in Table 2.

	Prop. Red	**Lacunarity (r=5)**	**Lacunarity (r=100)**	**Contagion**	**Patch Density**
Prop.(Flower)	1%	0%	0%	0%	4%
Prop.(Plant)	60%*	62%*	61%*	65%*	68%*
Prop.(Error)	38%	37%	38%	34%	27%

Table 2: Genetic variation between, not within plants, explains the observed phenotypic variation in all metrics. Results of the linear mixed model, described as percent of variance explained by each parameter.
* indicates significant variation explained by this component. Plants, and genotype by extension, are a key component controlling pattern formation, while flower contributes very little to phenotype. This suggests that the genetically identical flowers create very similar patterns, but individual plants with novel genotypes produce flowers with unique patterns.

Linear mixed models demonstrate that all of these metrics are primarily defined by the plant, and by extension, genotype. The variation in each metric is at least 60% explained between plants. In contrast, there is little to no observed variation between flowers of the same plant (Table 2). Phenotype, or pattern, is formed largely by a genetic component, Prop.(Plant), with individual flowers having little impact on phenotype. In short, patterns within a plant are similar, likely due to shared genes, but patterns are different between non-identical plants. Some variation, however, appears to be random (ε). This variation may arise from a multitude of reasons: nutrient uptake during development, infection of pathogens that impacts growth, etc.

Plants were grown in a common garden, so it is unlikely that environmental factors are the root of this unidentified variation. Ultimately, we conclude that there must be some genetic basis to pattern formation that contributes to the majority of the observed patterns. Further work will focus on identifying the underlying genetic mechanisms through which plants generate their patterns.

Segregation of F2 Population by Phenotype

In order to identify candidate genetic features that contribute to pattern formation, we must separate the population by pattern morphology. So far, we have used our protocol to characterize phenotypes quantitatively and discern differences between the patterns formed on top and bottom petals, or between plants. These data suggest that several genetic mechanisms are at work in regulating different aspects of pattern development. Now, we use the data collected from high throughput phenotyping to determine if our F2 population will cluster based on our phenotypic parameters. From data gathered in the linear mixed model, we expect to see flowers from the same plant to group together, but plants should separate from each other indiscriminately.

We plotted the principal components for top and bottom petals to see if the population separates by phenotype (Fig. 18). To provide a proof-of-concept, a small sample of eight plants, each with two to five flowers, is used to illustrate phenotypic separation. Plants are color- and symbol-coordinated for easy visualization. As expected, flowers from the same plant tend to cluster together. This is expected due to the genetic similarity of flowers from the same plant; identical genotype should lead to very similar phenotypes. Individual plants are dispersed across

the plot, generating two clear clusters. This plot, and this phenotyping approach as a whole, will be useful for separating a population prior to conducting a bulk segregant analysis to identify candidate patterning genes.

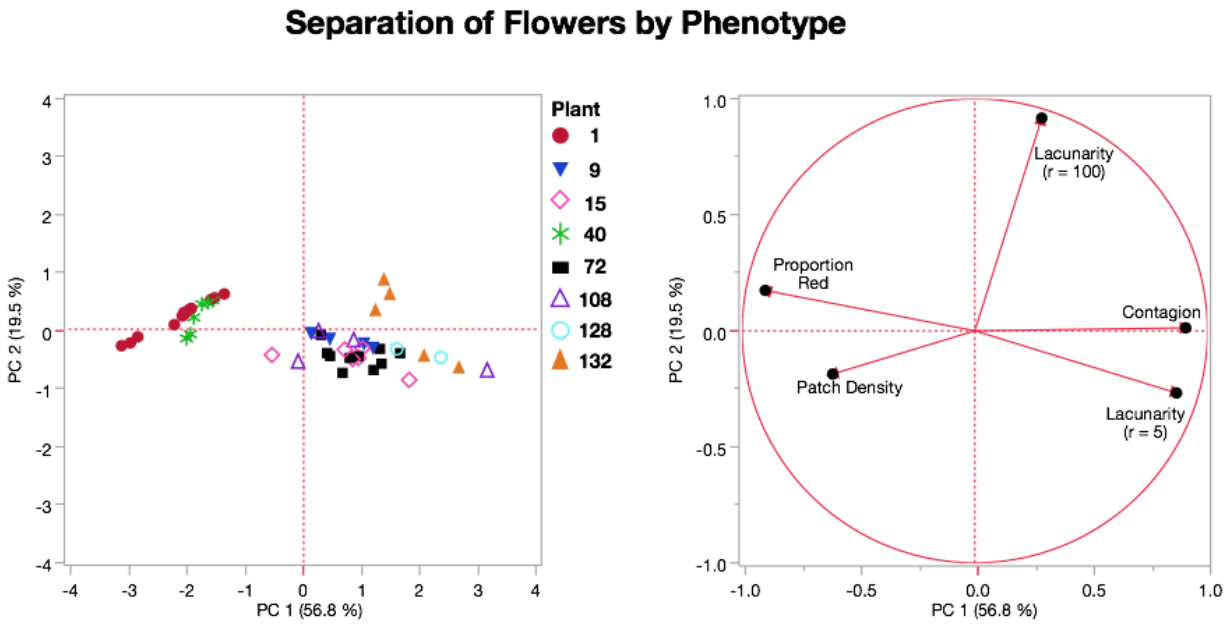

Figure 18: Flowers cluster by plant based on phenotyping data. 8 plants were plotted by principal components to determine if phenotyping procedures can separate by morphology. A small sample size is shown here for proof-of-concept. Flowers cluster by plant, but plants segregate into groups.

Identification of Candidate Regulatory Genes

Unfortunately, much of the genetic analysis remains to be done. Currently, DNA extracts from over 120 samples wait to be converted into libraries. During DNA library preparation, there are four key steps: restriction digest, adapter ligation, PCR amplification, and purification on magnetic beads. However, persistent issues with library preparation impede progress. Attempts to troubleshoot are discussed here.

We first investigate the restriction digest stage. There are several possible causes for digestion failure. One possibility is the presence of compounds that sequester cofactors essential

for restriction enzymes, such as EDTA. EDTA is often added to elution buffers to stabilize DNA for long periods of time; however, EDTA also binds $Mg2^{+}$, a metal ion is required for restriction enzyme function. EDTA contamination is unlikely to be the cause, as most if not all of the samples were eluted post-extraction with sterile deionized water. Another possibility for failure at the digest step is the lack of cut sites available for our specific enzyme, ApeKI. Again, this probably is not the case, as prior work has demonstrated that ApeKI efficiently cuts *Mimulus* DNA (Twyford and Friedman, 2015).

Alternatively, failure may occur at the ligation and amplification stages. Adapter ligation is dependent on ApeKI working effectively, as adapters contain sequences complementary to the sticky ends created at cut sites. It is possible that the adapters may be faulty, lacking a sequence that can base pair to the sticky ends. If the adapters fail to ligate, then all subsequent library steps would fail as well. Additionally, PCR could fail due to suboptimal annealing temperatures, preventing primers from binding and ultimately inhibiting DNA elongation.

The last, and most likely, possibility is that the DNA is lost during the AMPure bead clean-up. The DNA must air-dry for a few minutes so all ethanol from prior wash steps evaporates. Leaving samples out too long causes the DNA to dry to the beads and be lost, but inadequate drying times result in increased ethanol contamination. We believe that we have yet to optimize this step, and frequent DNA loss must be resolved.

In an attempt to identify the problematic stage, we attempted a series of bioanalyzer analyses at each stage of the protocol. Bioanalyzers use electric currents to estimate DNA fragment size and displays the results as a distribution curve (Fig. 19). Restriction digests appear to be successful, despite somewhat large fragment sizes. There is little evidence that any

amplification occurs during the PCR steps, as seen by the low concentration of fragments. In addition to this, erratic spiking in plots for what should be the final product indicates that ethanol contaminates the samples (not shown). These results suggest that much of the protocol needs further optimization, specifically at the PCR and clean-up steps.

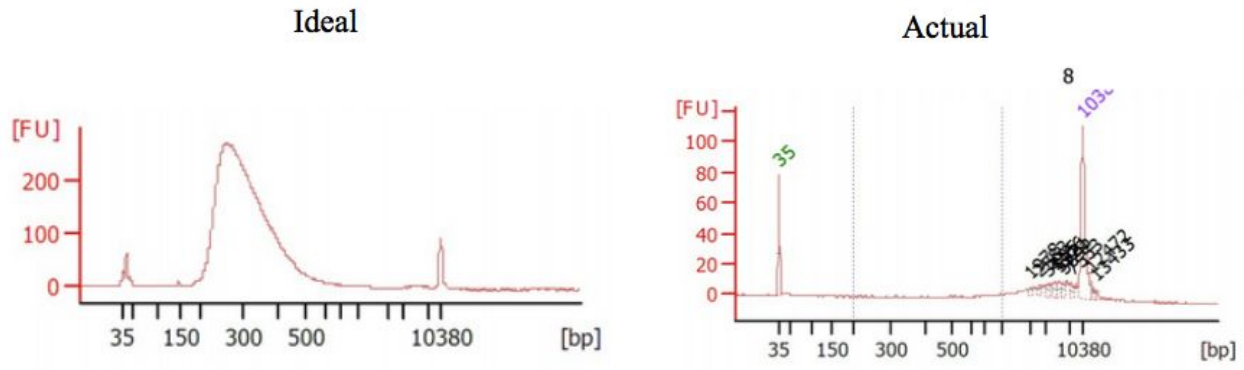

Figure 19: DNA libraries failed during preparation. Optimization of library preparation failed after twelve attempts. **Left**: An ideal DNA library, which a broad distribution of DNA fragments at a high concentration. X axis = fragment size (bps); Y axis = concentration (fluorescent units). Image courteously provided by Tal Kinser of Puzey Lab. **Right**: Attempted DNA library shows minimal fragmentation, with large post-digestion fragments, and low product concentration.

Discussion

A Genetic Basis for Pattern Formation

Little work has examined the genetic basis of pattern formation in plants. Color patterns are diverse across all forms of life, and the manner through which this diversity has evolved is unclear. We propose that pattern formation is a genetically regulated process of development, and that small changes to the genetic code leads to the dramatic divergence of patterns and the

creation of the diversity found in life. Most current research has focused on how patterns form during development in animals, such as mice or butterflies (Kunte et al., 2014; Mallarino et al., 2016). The insight from animal studies, however, may not translate directly to plants, due the early divergence of plants from the other kingdoms during the evolution of multicellularity. To increase the body of knowledge regarding genetically guided pattern formation in plants, we propose a framework for analyzing the phenotypic and genotypic factors that lead to pattern diversity, and using these factors to mathematically model how genetic changes lead to phenotypic diversity. Here, we present the first step towards building this framework: a high throughput phenotyping protocol designed to separate flower populations by color pattern morphology.

Our long-term goal is to provide an experimental design model for a holistic approach to genome-wide association studies through the implementation of high throughput, unbiased phenotyping, bulk segregant analysis, RAD-sequencing, and mathematical modeling. By applying all of these approaches, one can elucidate the whole picture of simple or complex genetics interactions that produce phenotypes during development. Ultimately, this approach will unite the areas of evolutionary genetics, developmental biology, and applied science to illustrate the role of genetics in pattern formation.

We have demonstrated that underneath the visual complexity of flower pattern development, there is a genetic basis guiding the formation of these patterns during development. We reach this conclusion based on two lines of evidence.

First, the top two petals of each flower are very similar in terms of patterning. If pattern formation were random, we would expect to see different patterns on each petal. Our

observations suggest that there is a genetic set of rules that guide pattern formation specifically for the top petals. In contrast, bottom petals appear to share some of these guidelines, but their basis for pattern formation is more complex, seen in their discrepancies from top petal patterns. It is possible that there are two separate mechanisms through which flowers regulate pattern formation in a location-dependent manner. While we did not assess the similarity between bottom left and right petals, visual observation shows that left and right bottom petals are also very similar to each other, but differ from the top. These petals likely develop patterns through a genetic mechanism similar to the top petals, but determining whether they follow all, some, or none of the same genetic rules as the top petals requires further experimentation.

Additional evidence for pattern-forming genes comes from the linear mixed model of our samples. We found that much of the observed pattern variation could be explained by genotype; each plant creates its own pattern for all of the flowers it produces, and flowers from the same plant all develop very similar patterns. If pattern formation were random, we would expect to see no specific parameter, such as genotype, wholly contributing to variation. The fact that we see most variation occurring between, but not within plants, indicates that the genes of each plant are responsible for pattern formation. There is still some unexplained variation between left and right petal patterns; we do not know if this is due to genetic differences, environmental effects, or error during experimentation. Further work is required to clarify the basis of this minor variation.

Based on the evidence found in this study, we are able to separate out plants by phenotype in an efficient and objective manner. Analysis of the metrics outlined here can be used for a principal component analysis, dividing the population into groups of similar phenotype.

Now that we have established that there is a genetic basis for pattern formation, we are confident in moving ahead towards mapping pattern-forming genes.

Pattern Formation Genetics and Evolutionary Impacts

These discoveries have several noteworthy implications. Heritable flower patterns may become subject to selection by pollinators, beginning the process of pollinator-dependent sympatric speciation in a local population. Pollinator preference can ultimately lead to prezygotic reproductive isolation, with further reinforcement of speciation over time (Schemske and Bradshaw, 1999). Here, pattern diversity could result in the formation of a novel species.

We also provide support to the hypothesis that the extreme diversity of color patterns, and flowering plants as a whole, arises through simple genetic changes to an existing framework. Small tweaks in the genetic code, such as the strength of a patterning gene's promoter, would impact the final pattern formed after development. If SNPs and alleles in a few loci can change pattern formation, evolution can rapidly influence a species' fitness through selection on various patterns and accumulation of mutations. This system could serve as a model for developmental biologists as they characterize complex trait coordination in organisms from plants to people. Here, we focus our investigations on plants, but this model may be extended to other organisms as well.

Applications Outside of Mimulus

Several of the metrics we used are applicable to other phenotyping projects as well. Contagion is useful for measuring patch aggregation. Because this metric takes each type of

patch into account, it would be useful for analysis of complex, non-binary color patterns. Proportion red and patch density are simple metrics that are applicable to most patterning studies. Lacunarity is a measure of randomness, but it is also a measure of rotational symmetry as well; a lower lacunarity value represents more conserved rotational symmetry. While this is not quite applicable in the case of single petal analysis, it may be worth investigating how lacunarity is calculated in whole-flower images to determine overall symmetry in the flower's pattern. Transferring this technique to other organisms may be useful for modeling circular patterns, such as sunflower floret organization or strawberry seed placement.

Further Protocol Optimization

Here, we provide a novel protocol for high throughput phenotype quantification, and we demonstrate that data gathered from this procedure can be used to separate a population into groups by phenotype. As such, it will have useful applications future studies involving bulk segregant analysis and pattern formation. Our protocol accelerates the characterization of phenotypes through a high throughput analysis using Adobe Photoshop and ArcGIS.

However, this protocol could be optimized to improve rates of phenotyping. While much of the phenotyping process is high throughput, some stages are time consuming. Notably, isolation of petals from raw images in Photoshop takes several hours if one attempts to do a large number of flowers at once. One possible alternative is the use of an automated program for petal isolation. ImageJ, image analysis software provided for free by the National Institute of Health for life science imaging, could substitute Photoshop, and automation is possible through the use of macros and scripts. However, one would need to explicitly define "petal" versus

"background" in the program, or else the software may inappropriately detect background pixels as "petal." The simplest alternative is to image each petal separately. This would accelerate not only the isolation from the raw image, but ArcGIS steps as well, as having a single petal per image nullifies the need for the Extract by Mask step.

Identifying Genes in Pattern Formation

We were unable to complete the genotyping portion of this project. To go on, we need to identify and remedy the roadblocks in GBS library preparation. From there, we will sequence the genomes of our sample population and determine which gene regions contribute the most to the observed phenotypes. After phenotypic grouping, DNA sequencing between groups will identify genes, alleles, or SNPs that are consistent within a group, but diverge when compared between groups due to F2 recombination.

To truly test if and how these genes influence pattern development, an RNA knockdown should be conducted. By transfecting a plasmid containing an RNAi construct into the germline of a plant, this plant's offspring will likely overexpress the genes from the start of development. Gene overexpression will induce RNA interference through the DICER/RISC pathway, and ultimately disrupt gene expression during development. Once the offspring fully grow, their phenotype can be characterized and we can see how the genes impact pattern generation in a wild-type flower. For example, knockdown of one gene may cause only the bottom center petal to be splotched with red, while the rest of the petals exhibit normal phenotypes; from this, we could conclude that the gene of interest contributes to spatial arrangement of pigment during development. As the promoters, activators and repressors are characterized for these genes, we

can begin to model how minor changes in the genetic variables, such as increased protomer affinity of a transcription factor, contributes to pattern development.

Closing Remarks

In the long-term, this experiment will hopefully serve as a model for evolutionary and developmental biologists seeking to characterize complex developmental traits from a holistic perspective. Already, we provide much of the foundation for the characterization of phenotype, and are now working to optimize rapid and cheap genotyping for large-scale experiments. Ideally, the entirety of this experiment will be completed within the next year. For that to happen, roughly 150 more flowers need to be quantified, and nearly the same number of DNA samples need to be extracted in order to reach the GWAS Goldilocks sample number of 300, the nexus of cost-effectiveness and high-resolution data. For now, however, the foundation is laid for all future work for flower pattern characterization.

Acknowledgements

This project would not be possible without the endless support of Dr. Josh Puzey. As a mentor, he has taught me most of what I know, from basic wet-lab procedures to the intricacies of scientific communications. I deeply appreciate the numerous hours that he contributed to helping me navigate this project. I would also like to thank Dr. Matthias Leu, who lent me use of his lab for ArcGIS and FragStats. His instruction with this software played a pivotal role in this project's success, and his enthusiasm encouraged me throughout the experiment. On the other side the country, I also thank Dr. Arielle Cooley of Whitman College for the creation of our F2 population, without which this project could have never begun. Her generosity is deeply appreciated. Across the Atlantic, I would like to thank Dr. Alex Twyford for providing the adapters necessary for DNA library preparation. In addition to the contributions of these professors, I also thank my Honors Committee, which consists of Dr. Puzey, Dr. Leu, Dr. Helen Murphy, and Dr. Jordan Walk. I am honored to have them on my committee and appreciate their time greatly.

Within the Puzey Lab, I would like to thank Jake Brammer and Hunter Call, who were always willing to assist with data analysis and computational work. On the botanical side, I thank Mary Bea Halloran and Scott Teresi, both of whom helped me with plant care and image analysis. Tal Kinser, a graduate student in Puzey Lab, was also always willing to help, even at 1:00am while in the midst of a cDNA library preparation. He always brings joy to the lab, and I appreciate all the help he gave me.

I also thank Michelle Brown, a student in Dr. Leu's lab who developed the preliminary protocol using ArcGIS, and her initial work was the inspiration for this project. Additionally, this thesis could not have completed without Wawa, a critical source of nutrients and caffeine that fueled the scientific and creative process from start to finish. Lastly, I would like to thank the Charles Center donors who funded this project, allowing me to kickstart the work over the summer of 2016 and all of this before graduation.

References

Allain, C., & Cloitre, M. (1991). Characterizing the lacunarity of random and deterministic fractal sets. *Physical Review A*, *44*(6), 3552–3558. https://doi.org/10.1103/PhysRevA.44.3552

Ball, P. (2013). In retrospect: On Growth and Form. *Nature*, *494*(7435), 32–33. https://doi.org/10.1038/494032a

De Bodt, S., Maere, S., & Van de Peer, Y. (2005). Genome duplication and the origin of angiosperms. *Trends in Ecology & Evolution*, *20*(11), 591–597. https://doi.org/10.1016/j.tree.2005.07.008

Clarke, C. A., & Sheppard, P. M. (1972). The Genetics of the Mimetic Butterfly Papilio polytes L. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *263*(855), 431–458. https://doi.org/10.1098/rstb.1972.0006

Cooley, A. M., & Willis, J. H. (2009). Genetic divergence causes parallel evolution of flower color in Chilean Mimulus. *New Phytologist*, *183*(3), 729–739. https://doi.org/10.1111/j.1469-8137.2009.02858.x

Cordell, H. J., & Clayton, D. G. (2005). Genetic association studies. *The Lancet*, *366*(9491), 1121–1131. https://doi.org/10.1016/S0140-6736(05)67424-7

Ding, B., & Yuan, Y.-W. (2016). Testing the utility of fluorescent proteins in Mimulus lewisii by an Agrobacterium-mediated transient assay. *Plant Cell Reports*, *35*(4), 771–777. https://doi.org/10.1007/s00299-015-1919-1

Edwards, L. (2010, September 23). Estimate of flowering plant species to be cut by 600,000. Retrieved April 27, 2017, from https://phys.org/news/2010-09-species.html

Greer, P. L., & Greenberg, M. E. (2008). From Synapse to Nucleus: Calcium-Dependent Gene Transcription in the Control of Synapse Development and Function. *Neuron*, *59*(6), 846–860. https://doi.org/10.1016/j.neuron.2008.09.002

Iwashina, T. (2000). The Structure and Distribution of the Flavonoids in Plants. *Journal of Plant Research*, *113*(3), 287–299. https://doi.org/10.1007/PL00013940

Jaakola, L., Määttä, K., Pirttilä, A. M., Törrönen, R., Kärenlampi, S., & Hohtola, A. (2002). Expression of Genes Involved in Anthocyanin Biosynthesis in Relation to Anthocyanin, Proanthocyanidin, and Flavonol Levels during Bilberry Fruit Development. *Plant Physiology*, *130*(2), 729–739. https://doi.org/10.1104/pp.006957

Knoll, A. H. (2011). The Multiple Origins of Complex Multicellularity. *Annual Review of Earth and Planetary Sciences*, *39*(1), 217–239. https://doi.org/10.1146/annurev.earth.031208.100209

Koes, R., Verweij, W., & Quattrocchio, F. (2005). Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. *Trends in Plant Science*, *10*(5), 236–242. https://doi.org/10.1016/j.tplants.2005.03.002

Kunte, K., Zhang, W., Tenger-Trolander, A., Palmer, D. H., Martin, A., Reed, R. D., … Kronforst, M. R. (2014). doublesex is a mimicry supergene. *Nature*, *507*(7491), 229–232. https://doi.org/10.1038/nature13112

Mallarino, R., Henegar, C., Mirasierra, M., Manceau, M., Schradin, C., Vallejo, M., … Hoekstra, H. E. (2016). Developmental mechanisms of stripe patterns in rodents. *Nature*, *539*(7630), 518–523. https://doi.org/10.1038/nature20109

McGarigal, K., SA Cushman, and E Ene. 2012. FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical and Continuous Maps. Computer software program produced by the authors at

the University of Massachusetts, Amherst. Available at the following web site: http://www.umass.edu/landeco/research/fragstats/fragstats.html

Medel, R., Botto-Mahan, C., & Kalin-Arroyo, M. (2003). POLLINATOR-MEDIATED SELECTION ON THE NECTAR GUIDE PHENOTYPE IN THE ANDEAN MONKEY FLOWER, MIMULUS LUTEUS. *Ecology*, *84*(7), 1721–1732. https://doi.org/10.1890/01-0688

Müller, G. B. (2007). Evo–devo: extending the evolutionary synthesis. *Nature Reviews Genetics*, *8*(12), 943–949. https://doi.org/10.1038/nrg2219

Nadeau, J. A., & Sack, F. D. (2002). Control of Stomatal Distribution on the Arabidopsis Leaf Surface. *Science*, *296*(5573), 1697–1700. https://doi.org/10.1126/science.1069596

Nguyen, V. H., Schmid, B., Trout, J., Connors, S. A., Ekker, M., & Mullins, M. C. (1998). Ventral and Lateral Regions of the Zebrafish Gastrula, Including the Neural Crest Progenitors, Are Established by abmp2b/swirlPathway of Genes. *Developmental Biology*, *199*(1), 93–110. https://doi.org/10.1006/dbio.1998.8927

Pattern. (1999) In *Merriam-Webster's Collegiate Dictionary*. Retrieved from https://www.merriam-webster.com/dictionary/pattern.

Peterson, E. E., & Hoef, J. M. V. (2010). A mixed-model moving-average approach to geostatistical modeling in stream networks. *Ecology*, *91*(3), 644–651. https://doi.org/10.1890/08-1668.1

Plotnick, R. E., Gardner, R. H., Hargrove, W. W., Prestegaard, K., & Perlmutter, M. (1996). Lacunarity analysis: A general technique for the analysis of spatial patterns. *Physical Review E*, *53*(5), 5461–5468. https://doi.org/10.1103/PhysRevE.53.5461

Reinhardt, D., Pesce, E.-R., Stieger, P., Mandel, T., Baltensperger, K., Bennett, M., … Kuhlemeier, C. (2003). Regulation of phyllotaxis by polar auxin transport. *Nature*, *426*(6964), 255–260. https://doi.org/10.1038/nature02081

Schemske, D. W., & Bradshaw, H. D. (1999). Pollinator preference and the evolution of floral traits in monkeyflowers (Mimulus). *Proceedings of the National Academy of Sciences*, *96*(21), 11910–11915. https://doi.org/10.1073/pnas.96.21.11910

Simbolo, M., Gottardi, M., Corbo, V., Fassan, M., Mafficini, A., Malpeli, G., … Scarpa, A. (2013). DNA Qualification Workflow for Next Generation Sequencing of Histopathological Samples. *PLOS ONE*, *8*(6), e62692. https://doi.org/10.1371/journal.pone.0062692

Smith, R. S., Guyomarc'h, S., Mandel, T., Reinhardt, D., Kuhlemeier, C., & Prusinkiewicz, P. (2006). A plausible model of phyllotaxis. *Proceedings of the National Academy of Sciences*, *103*(5), 1301–1306. https://doi.org/10.1073/pnas.0510457103

Stoehr, A. M., Walker, J. F., & Monteiro, A. (2013). Spalt expression and the development of melanic color patterns in pierid butterflies. *EvoDevo*, *4*, 6. https://doi.org/10.1186/2041-9139-4-6

Suzuki, K., Xue, H., Tanaka, Y., Fukui, Y., Fukuchi-Mizutani, M., Murakami, Y., … Kusumi, T. (2000). Flower color modifications of Torenia hybrida by cosuppression of anthocyanin biosynthesis genes. *Molecular Breeding*, *6*(3), 239–246. https://doi.org/10.1023/A:1009678514695

Twyford, A. D., & Friedman, J. (2015). Adaptive divergence in the monkey flower Mimulus guttatus is maintained by a chromosomal inversion. *Evolution*, *69*(6), 1476–1486. https://doi.org/10.1111/evo.12663

"Who Uses ArcGIS Server?" ESRI 2011. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute.

Wu, C. A., Lowry, D. B., Cooley, A. M., Wright, K. M., Lee, Y. W., & Willis, J. H. (2008). Mimulus is an emerging model system for the integration of ecological and genomic studies. *Heredity*, *100*(2), 220–230. https://doi.org/10.1038/sj.hdy.6801018

Yuan, Y.-W., Rebocho, A. B., Sagawa, J. M., Stanley, L. E., & Bradshaw, H. D. (2016). Competition between anthocyanin and flavonol biosynthesis produces spatial pattern variation of floral pigments between *Mimulus* species. *Proceedings of the National Academy of Sciences*, *113*(9), 2448–2453. https://doi.org/10.1073/pnas.1515294113

Yuan, Y.-W., Sagawa, J. M., Frost, L., Vela, J. P., & Bradshaw, H. D. (2014). Transcriptional control of floral anthocyanin pigmentation in monkeyflowers (*Mimulus*). *New Phytologist*, *204*(4), 1013–1027. https://doi.org/10.1111/nph.12968